



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2021

A survey of preference estimation with unobserved choice set heterogeneity

Crawford, Gregory S ; Griffith, Rachel ; Iaria, Alessandro

Abstract: We provide an introduction to the estimation of discrete choice models when choice sets are heterogeneous and unobserved to the econometrician. We survey the two most popular approaches: “integrating over” and “differencing out” unobserved choice sets. Inspired by Chamberlain (1980)’s original idea of constructing sufficient statistics from observed choices, we introduce the term “sufficient set” to refer to any combination of observed choices that lies within the true but unobserved choice set. The concept of sufficient set helps to unify notation and organize our thinking, to map econometric assumptions onto economic models, and to implement both methods in practice.

DOI: <https://doi.org/10.1016/j.jeconom.2020.07.024>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-190402>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.

Originally published at:

Crawford, Gregory S; Griffith, Rachel; Iaria, Alessandro (2021). A survey of preference estimation with unobserved choice set heterogeneity. *Journal of Econometrics*, 222(1):4-43.

DOI: <https://doi.org/10.1016/j.jeconom.2020.07.024>



Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A survey of preference estimation with unobserved choice set heterogeneity[☆]

Gregory S. Crawford^{a,b}, Rachel Griffith^{b,c,d}, Alessandro Iaria^{b,e,*}

^a Department of Economics, University of Zürich, Switzerland

^b CEPR, United Kingdom of Great Britain and Northern Ireland

^c Department of Economics, University of Manchester, United Kingdom of Great Britain and Northern Ireland

^d IFS, United Kingdom of Great Britain and Northern Ireland

^e Department of Economics, University of Bristol, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Article history:

Available online xxxx

JEL classification:

C18
C33
C35
D12
D83

Keywords:

Discrete choice models
Panel data
Unobserved choice sets
Unobserved heterogeneity
Sufficient sets

ABSTRACT

We provide an introduction to the estimation of discrete choice models when choice sets are heterogeneous and unobserved to the econometrician. We survey the two most popular approaches: “integrating over” and “differencing out” unobserved choice sets. Inspired by Chamberlain (1980)’s original idea of constructing sufficient statistics from observed choices, we introduce the term “sufficient set” to refer to any combination of observed choices that lies within the true but unobserved choice set. The concept of sufficient set helps to unify notation and organize our thinking, to map econometric assumptions onto economic models, and to implement both methods in practice.

© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Discrete choice models are commonly used in applied economics. To estimate a discrete choice model requires that we have information on the set of alternatives over which the consumer is choosing. In many situations it is trivial to observe consumers’ choice sets, for example, Bay Area commuters’ mode choice was between car and bus before BART was built. In other cases, researchers can explicitly ask survey respondents about the alternatives that they considered or ask for a ranking (e.g., Train and Winston, 2007, Berry et al., 2004). However, in many situations it can be difficult to formulate choice sets, for example, the discrete choice of housing with the constraint that there is only one house for sale at the specific address the buyer prefers (de Palma et al., 2007). There are many reasons why consumer choice sets may be heterogeneous and unobserved by econometricians, including limited consumer attention, search, or endogenous product choices by firms. Failing to account for unobserved choice set heterogeneity will generally cause estimators of preference parameters to be inconsistent.

[☆] Griffith gratefully acknowledges financial support from the European Research Council (ERC) under ERC-2009-AdG-249529 and ERC-20150AdG-694822, and the Economic and Social Research Council (ESRC) under RES-544-28-0001 and ES/N011562/1. Iaria gratefully acknowledges financial support from the research grant Labex Ecodec: Investissements d’Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047).

* Corresponding author at: Department of Economics, University of Bristol, United Kingdom of Great Britain and Northern Ireland.

E-mail addresses: gregory.crawford@econ.uzh.ch (G.S. Crawford), rgriffith@ifs.org.uk (R. Griffith), alessandro.iaria@bristol.ac.uk (A. Iaria).

<https://doi.org/10.1016/j.jeconom.2020.07.024>

0304-4076/© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

In this paper, we survey the two main empirical approaches to tackling the problem of unobserved choice set heterogeneity: “integrating over” and “differencing out” unobserved choice sets. The two approaches originate from different econometric literatures, started respectively by [Manski \(1977\)](#) and [McFadden \(1978\)](#). While integrating over heterogeneous unobserved choice sets is commonly done in empirical applications, differencing them out is less popular, possibly because [McFadden \(1978\)](#)’s original motivation was to facilitate estimation with large but *observed* choice sets. We provide a unifying notation for understanding the two approaches and, inspired by [Chamberlain \(1980\)](#)’s idea of constructing sufficient statistics from observed choices, we introduce the idea of “sufficient sets”, which serve several purposes. First, sufficient sets help clarify that differencing out can also address the problem of unobserved choice sets, and that it is complementary to integrating over them. Second, sufficient sets prove useful to implement both approaches in practice, particularly in panel-data environments. Third, they help translate economic assumptions derived from the characteristics of a given choice environment into econometric assumptions appropriate for estimation.

To build intuition, we begin by illustrating the well-understood result that, in a multinomial logit (MNL), mistakenly adding alternatives to a consumer’s choice set leads to violations of the Independence of Irrelevant Alternatives (IIA) and to inconsistent estimators. This problem arises naturally with unobserved choice set heterogeneity when researchers specify a common choice set for all consumers that is likely to be larger than at least some consumers’ true choice sets, for example by specifying a choice set that includes all products above a certain market share threshold or a given number of products with the largest product shares.

The most popular approach to address this concern models the joint probability that a consumer is matched to a certain choice set and that she makes a specific choice from that choice set. As first proposed by [Manski \(1977\)](#), one can then obtain the marginal choice probability by “integrating over” unobserved choice set heterogeneity, as is routinely done with any other form of unobserved heterogeneity (e.g., standard mixed logit models that integrate over unobserved preference heterogeneity). The integrating over approach has a long tradition in marketing and transportation studies analyzing consumers’ consideration sets (e.g., [Roberts and Lattin, 1991](#), [Ben-Akiva and Boccara, 1995](#), [Bronnenberg and Vanhonacker, 1996](#), [Chiang et al., 1998](#), [Başar and Bhat, 2004](#), [Erdem and Swait, 2004](#), [Bruno and Vilcassim, 2008](#), [Van Nierop et al., 2010](#), [Draganska and Klapper, 2011](#), [Ching et al., 2014](#), and [Juang and Bronnenberg, 2018](#)) and has also recently been applied in economics (e.g., [Goeree, 2008](#), [De los Santos et al., 2012](#), [Conlon and Mortimer, 2013](#), [Honka, 2014](#), [Abaluck and Adams, 2017](#)).¹ Because the number of choice sets to integrate over grows exponentially in the number of available products, practical implementation is subject to a curse of dimensionality.

The insight that mistakenly adding alternatives to choice sets yields inconsistent estimates also forms the basis of the “differencing out” approach proposed by [McFadden \(1978\)](#) for the consistent estimation of MNL models from subsets of true and observed choice sets. The idea of estimating discrete choice models from subsets of true/observed choice sets has been shown to be effective also in Generalized Extreme Value models (e.g., [Train et al., 1987](#), [Bierlaire et al., 2008](#), and [Guevara and Ben-Akiva, 2013b](#)), mixed logit models with discrete distributions of random coefficients (e.g., [Bajari et al., 2007](#), [Fox et al., 2011](#), and [Fox et al., 2016](#)), semi-parametric models (e.g., [Fox, 2007](#)) and, in the context of “long” panel data, individual-specific MNL models (e.g., [Dubois et al., 2020](#)).²

However, it may not be clear to a researcher how to construct proper subsets of consumers’ true choice sets when they are *unobserved*, and therefore how to implement these estimators (e.g., [Frejinger et al., 2009](#)). Inspired by [Chamberlain \(1980\)](#), we propose the use of consumers’ observed choices paired with assumptions about the evolution of their unobserved choice sets over time as a practical tool to construct proper subsets in panel data environments characterized by unobserved choice set heterogeneity. We call these subsets “sufficient sets”. For example, in the case of the MNL with fixed effects studied by [Chamberlain \(1980\)](#), the set of permutations of a consumer’s observed choices over time represents a sufficient set, indeed one which ([Chamberlain, 1980](#)) specifically chose to difference out fixed effects but that incidentally also differences out true and potentially unobserved choice sets.

We expand on this idea in a wide variety of models using the “differencing out” approach ranging from the original MNL model of [McFadden \(1978\)](#) to the semi-parametric Pairwise Maximum Score Estimator of [Fox \(2007\)](#). We also show that sufficient sets can be used in the “integrating over” approach to greatly reduce the number of choice sets over which a researcher needs to integrate (i.e., only *supersets* of the sufficient sets can have positive probability mass), making this approach more viable in applications with even moderately large choice sets.³

We highlight the costs and benefits of differencing out unobserved choice sets relatively to integrating over them. The integrating over approach requires the specification of a choice set formation model which is estimated along with preferences given choice sets, thus requiring additional functional form assumptions, data on the choice set formation process, and is computationally more intensive. However, it enables researchers to learn about both consumer preferences and choice set formation. This may be essential in applications in which the key counterfactuals involve re-matching of choice sets to consumers (e.g., [Gaynor et al., 2016](#)). The differencing out approach treats the choice set formation

¹ For a recent survey of these applied literatures in the context of consumer search, see [Honka et al. \(2019\)](#).

² To the best of our knowledge, in the context of cross-sectional or “short” panel data, results of this kind are not available for mixed logit models with continuous distributions of random coefficients, even though some interesting approximations have been proposed by [Keane and Wasi \(2012\)](#) and [Guevara and Ben-Akiva \(2013a\)](#).

³ [Goeree \(2008\)](#) proposes a convenient importance sampling procedure (which we detail in [Appendix A](#)) that also greatly simplifies the practical implementation of this approach. [Goeree \(2008\)](#)’s importance sampling and sufficient sets are not mutually exclusive and can be used together.

process as a nuisance parameter to be dropped from the likelihood function, so requires less prior knowledge and data on the choice set formation process, and it is simpler to implement. However, it does not allow inference about choice set formation. Whenever information on the choice set formation process is available, integrating over is likely to be the preferred approach. However, there are cases where such information is not available or is unreliable, or where the curse of dimensionality prevents implementation of the integrating over approach. In these cases, one can still learn about consumer preferences by differencing out unobserved choice sets.

Importantly, the economic characteristics of the choice environment can inform the specification of sufficient sets in both approaches. Sufficient sets can help map concrete economic information about choice environments onto suitable econometric assumptions. For example, settings characterized by non-sequential or fixed-sample search (e.g., [Morgan and Manning, 1985](#)) imply a choice environment that is stable for any individual consumer over time, but (possibly) varying across consumers. These models imply sufficient sets based on the collection of products chosen by a consumer over time, what we call a Full Purchase History sufficient set. Settings characterized by sequential search (e.g., [Caplin et al., 2011](#)), limited attention (e.g., [Eliaz and Spiegel, 2011](#)), or consumer focus (e.g., [Kőszegi and Szeidl, 2013](#)) imply choice environments that evolve over time, and suggest sufficient sets based on the accumulation of products chosen by a consumer in the past, what we call a Past Purchase History sufficient set. Cross-sectional settings where a group of individuals face a common choice environment, as for example in the analyses of [Currie et al. \(2010\)](#) of whether greater availability of fast food increases obesity, suggest sufficient sets made of the collection of products observed to be chosen by individuals in the relevant group, what we call an Inter-Personal sufficient set. These are only a few examples of sufficient sets. Sufficient sets can be combined or devised to reflect a large range of choice environments. We discuss the use of specification tests to aid comparisons between alternative sufficient sets within a particular application.

We implement some of the surveyed approaches in both Monte Carlo simulations and an empirical illustration. The illustration estimates the distribution of price and advertising sensitivities for chocolate bars; advertising can affect both a consumer's valuation of a specific product and the likelihood that the product enters the consumer's choice set. We show that assumptions on individuals' choice sets can have a material impact on the estimated distributions of price and advertising sensitivities. These results are in line with models of imperfect consumer attention as in [Eliaz and Spiegel \(2011\)](#) and the patterns of estimated price sensitivities are in line with those by [Goeree \(2008\)](#).

The methods surveyed in this paper rely on an exogeneity assumption between consumer preferences and the matching of consumers to choice sets. This assumption is ubiquitous in empirical work and accommodates models in which firms select the products to sell or in which consumers search for alternatives on the basis of both observable characteristics and expectations over unobservable characteristics, but rules out the matching of consumers to choice sets on the basis of the realizations of unobserved preferences.⁴ The problems of "endogenous" matching and of "unobserved" choice sets are of a fundamentally different nature and can be better understood in isolation. On the one hand, endogenous matching of choice sets gives rise to econometric problems akin to sample selection even when choice sets are perfectly observable.⁵ On the other, unobserved choice set heterogeneity is a concern even when choice sets are exogenously matched to consumers. In this paper we limit our discussion to exogenous unobserved choice set heterogeneity and refer the reader to [Hickman and Mortimer \(2016\)](#) and [Honka et al. \(2019\)](#) for recent surveys on the endogenous matching of choice sets.⁶ Some recent working papers, such as [Lu \(2018\)](#) and [Barseghyan et al. \(2019\)](#), address these problems on the basis of partial identification methods and illustrate that many interesting features of the model can be identified even when unobserved choice sets are endogenously matched to individuals on the basis of their unobserved preferences.

In addition to the empirical literatures surveyed above, this paper relates to a fast-growing theoretical literature in which limited attention is used to rationalize apparently inconsistent consumer and firm behaviors. This strongly motivates our interest in empirical approaches to accommodate such theories. These include, for example, consumer attention as in [Eliaz and Spiegel \(2011\)](#), [Masatlioglu et al. \(2012\)](#), and [Manzini and Mariotti \(2014\)](#), rational inattention as in [Gabaix \(2014\)](#) and [Matejka and McKay \(2015\)](#), search as in [Janssen and Moraga-González \(2004\)](#) and [Rhodes \(2014\)](#), screening rules as in [Gilbride and Allenby \(2004\)](#), models of salience as in [Bordalo et al. \(2014\)](#), and focus as in [Kőszegi and Szeidl \(2013\)](#).

The rest of the paper is structured as follows. In Section 2, we introduce the model and illustrate how heterogeneity in unobserved choice sets may lead to inconsistent estimators of MNL models. This helps to build intuition for the possible solutions discussed in Section 3, where we also (briefly) discuss the case of aggregate-level data on market shares. In Section 4 we discuss a number of other considerations in the differencing out approach, including the relationship between some economic models and sufficient sets, how to create bounds of functions of the point-identified preference parameters (e.g. price elasticities and consumer surplus), and specification tests. We provide an empirical illustration in Section 5. A final section concludes. Several appendices present additional derivations and results.

⁴ Most applied papers dealing with choice set formation processes also rely on assumptions that guarantee exogenous matching, see for example: [Goeree \(2008\)](#), [Draganska et al. \(2009\)](#), [Conlon and Mortimer \(2013\)](#), and [Eizenberg \(2014\)](#).

⁵ Some papers dealing with observed but endogenous choice set heterogeneity include [Iaria \(2014\)](#), [Musalem \(2015\)](#), [Ciliberto et al. \(2016\)](#), and [Li et al. \(2018\)](#).

⁶ In addition, [Hickman and Mortimer \(2016\)](#) and [Honka et al. \(2019\)](#) also discuss the problem of unobserved choice set heterogeneity in the context of aggregate-level data on market shares (e.g., [Goeree, 2008](#) and [Bruno and Vilcassim, 2008](#)). While we briefly mention this problem in the current paper, most of our discussion focuses on individual-level purchase data. In our view, the idea of "sufficient set" does not suit well the case of aggregate data: similar to [Chamberlain \(1980\)](#), we propose to construct subsets of each individual's unobserved choice set on the basis of observed individual-level past purchases. In the case of aggregate-level data, it is not clear how one could use panel data on the evolution of market shares to obtain similar results to those available for individual-level discrete choice models.

2. The problem of unobserved choice set heterogeneity

To build intuition, we start by illustrating that unobserved choice set heterogeneity leads to violations of the Independence of Irrelevant Alternatives (IIA) property in models with Gumbel errors and therefore to inconsistent estimators. Understanding the nature of the problem introduced by unobserved choice set heterogeneity helps to better appreciate the solutions proposed in the literature, which we will describe in the next section (including those that accommodate more general error specifications).

2.1. Basic model and notation

Let there be a panel of $i = 1, \dots, I$ individuals each observed making a sequence of T choices, one per choice situation $t = 1, \dots, T$. Denote i 's sequence of choices by $Y_i = (Y_{i1}, \dots, Y_{iT})$. For simplicity, we assume to observe exactly T choice situations for each i , but this is easily generalized.

We consider a situation in which i is matched to her choice set CS_{it}^* in choice situation t , but this choice set is unobserved to the researcher. Denote by \times the cartesian product and let i 's set of possible choice sequences be given by $CS_i^* = \times_{t=1}^T CS_{it}^*$. Any choice sequence Y_i belongs to CS_i^* .

Let preferences be defined by a parameter vector θ and let the probability with which i is matched to a given set of possible choice sequences, $CS_i^* = c$, be given by $\Pr[CS_i^* = c | \gamma]$, with γ also a parameter vector. In principle, γ could include some or all of the parameters that are in θ and could be the result of, for example, limited consumer attention, consumer search behavior, or strategic decision-making by firms.

Given θ and a specific match with a set of possible choice sequences, $CS_i^* = c$, each individual i is observed to make a sequence of choices $Y_i = j = (j_1, \dots, j_T)$. We assume that the conditional indirect utility of alternative j_t in choice situation t for individual i is

$$U_{ijt} = V(X_{ijt}, \theta) + \epsilon_{ijt}, \quad (2.1)$$

where X_{ijt} is a vector of observable characteristics, and ϵ_{ijt} is the portion of i 's utility that is unobserved to the econometrician. The probability that i is matched to the set of possible choice sequences $CS_i^* = c$ and makes a sequence of choices $Y_i = j$ is:

$$\Pr[Y_i = j, CS_i^* = c | \theta, \gamma] = \Pr[Y_i = j | CS_i^* = c, \theta] \Pr[CS_i^* = c | \gamma]. \quad (2.2)$$

The first term in (2.2) is the probability of choosing j solely due to preferences given the sequence of choice sets i is matched to, and the second is the probability that the individual is matched to that sequence of choice sets. The following assumption implies that $\Pr[Y_i = j | CS_i^* = c, \theta]$ is multinomial logit (MNL) for any c .

Assumption 1. Conditional on all $V(X_{ijt}, \theta)$'s and on $CS_i^* = c$, ϵ_{ijt} from (2.1) is distributed i.i.d. Gumbel.

Assumption 1 allows for general matching processes: $\Pr[CS_i^* = c | \gamma]$ can take any form and be a function of any element of the MNL model $\Pr[Y_i = j | CS_i^* = c, \theta]$. For example, this accommodates models in which firms select the products to sell or in which individuals search for alternatives on the basis of both observable characteristics and expectations over unobservable characteristics, but rules out the matching of individuals to choice sets on the basis of the *realizations* of the unobservables, ϵ_{ijt} 's. In other words, **Assumption 1** rules out the possibility that individuals and choice sets are *endogenously* matched.⁷

We use **Assumption 1** to provide some intuition about the econometric problem introduced by unobserved choice set heterogeneity. In Section 3 we use this to illustrate the basic ideas behind the main solutions proposed in the literature, we then relax **Assumption 1** and extend the discussion to more general models than the MNL.

An implication of **Assumption 1** is that conditional Maximum Likelihood estimators of θ can be constructed from $\Pr[Y_i = j | CS_i^* = c, \theta]$, since this conditional probability is multinomial logit.⁸ Using **Assumption 1**, i 's conditional probability of selecting choice sequence $Y_i = j$, given their set of possible choice sequences, $CS_i^* = c = \times_{t=1}^T c_t$, is the familiar product of T MNL's, each one specific to a choice situation along the sequence:

$$\Pr[Y_i = j | CS_i^* = c, \theta] = \prod_{t=1}^T \frac{\exp(V(X_{ijt}, \theta))}{\sum_{k \in CS_{it}^* = c_t} \exp(V(X_{ikt}, \theta))}. \quad (2.3)$$

⁷ See footnote 5 in the introduction for references on the observed but endogenous matching of choice sets in demand estimation.

⁸ Given **Assumption 1**, if γ shares some common element with θ (as is likely), failing to control for the choice set matching process $\Pr[CS_i^* = c | \gamma]$ only causes a loss of efficiency in the resulting conditional Maximum Likelihood estimator relative to a joint Maximum Likelihood estimator derived from Eq. (2.2).

2.2. The problem of adding unavailable alternatives

Eq. (2.3) cannot be directly used for estimation, because $CS_i^* = c$ is unobserved. In order to proceed, usually the researcher will specify a set of choice sequences, $S_i = s$, possibly different from $CS_i^* = c$, on the basis of which to construct a likelihood function. Researchers often specify S_i to be common across i and given by, for example, all those products above a certain market share threshold or a given number of products with the largest market shares. Suppose that the researcher specifies the likelihood function to be used in estimation as the conditional probability of i choosing $Y_i = j$ from the set of choice sequences $S_i = s = \times_{t=1}^T S_t$. Then, the potentially misspecified model is:

$$\Pr[Y_i = j | S_i = s, \theta] = \prod_{t=1}^T \frac{\exp(V(X_{ijt}, \theta))}{\sum_{k \in S_{it}=s_t} \exp(V(X_{ikt}, \theta))}, \quad (2.4)$$

where the difference between (2.3) and (2.4) lies in the terms included in the summations in the denominator of each. Since McFadden (1978), is known that using model (2.4) to estimate θ in the presence of unobserved choice set heterogeneity will *not* be a problem whenever the researcher manages to specify $S_i = s \subseteq CS_i^* = c$. However, whenever $S_i = s$ includes alternatives not originally available in $CS_i^* = c$, then the use of model (2.4) will lead to inconsistent estimators of θ . Not surprisingly, this is a consequence of model (2.4) satisfying (when $S_i = s \subseteq CS_i^* = c$) or violating (when $S_i = s \not\subseteq CS_i^* = c$) the IIA property from Assumption 1. To see this, note that individual i 's probability of choosing sequence j among the potentially misspecified S_i , given that the true set of sequences is $CS_i^* = c$ and conditional on the vector of parameters θ is:

$$\begin{aligned} \Pr[Y_i = j | S_i = s, CS_i^* = c, \theta] &= \prod_{t=1}^T \frac{\Pr[Y_{it} = j_t | CS_{it}^* = c_t, \theta]}{\sum_{r_t \in S_{it} \cap c_t} \Pr[Y_{it} = r_t | CS_{it}^* = c_t, \theta] + \sum_{k_t \in S_{it} \setminus c_t} \Pr[Y_{it} = k_t | CS_{it}^* = c_t, \theta]} \\ &= \prod_{t=1}^T \frac{\exp(V(X_{ijt}, \theta))}{\sum_{r_t \in S_{it}=s_t \cap CS_{it}^*=c_t} \exp(V(X_{irt}, \theta))} \\ &= \Pr[Y_i = j | S_i = s \cap CS_i^* = c, \theta], \end{aligned} \quad (2.5)$$

where the denominator in the first line decomposes s_t into those alternatives that are in c_t ($r_t \in s_t \cap c_t$) and those that are not ($k_t \in s_t \setminus c_t$). The second equality obtains because the probability i selects an alternative not in her true choice set is zero, $\Pr[Y_{it} = k_t | CS_{it}^* = c_t, \theta] = 0$ for all $k_t \notin CS_{it}^* = c_t$. In other words, because Assumption 1 implies the IIA property only when the choice set assumed by the researcher is a *subset* of i 's true choice set, $S_i = s \subseteq CS_i^* = c$, Eq. (2.5) is not guaranteed to equal (2.4). By expressing (2.5) in terms of (2.4), we obtain:

$$\begin{aligned} \Pr[Y_i = j | S_i = s \cap CS_i^* = c, \theta] &= \prod_{t=1}^T \frac{\exp(V(X_{ijt}, \theta))}{\sum_{r_t \in S_{it} \cap c_t} \exp(V(X_{irt}, \theta))} \frac{\sum_{m_t \in s_t} \exp(V(X_{imt}, \theta))}{\sum_{m_t \in s_t} \exp(V(X_{imt}, \theta))} \\ &= \prod_{t=1}^T \frac{\exp(V(X_{ijt}, \theta) - \ln\left(\frac{\sum_{r_t \in S_{it} \cap c_t} \exp(V(X_{irt}, \theta))}{\sum_{m_t \in s_t} \exp(V(X_{imt}, \theta))}\right))}{\sum_{m_t \in S_{it}=s_t} \exp(V(X_{imt}, \theta))} \\ &= \prod_{t=1}^T \frac{\exp(V(X_{ijt}, \theta) - \ln(\pi_{it}(\theta)))}{\sum_{m_t \in S_{it}=s_t} \exp(V(X_{imt}, \theta))}. \end{aligned} \quad (2.6)$$

Suppose $s_t \cap c_t \subset s_t$ for some t 's (i.e., s_t includes alternatives not in c_t), then $\ln(\pi_{it}) < 0$ for those t 's and models (2.4) and (2.5) will differ. In this case, if estimation proceeds on the basis of model (2.4), the likelihood function will be mistakenly ignoring a sequence of up to T fixed effects for each i , $\ln(\pi_{it})$'s, which are functions of the rest of the model. Put differently, suppose instead $s_t \subseteq c_t$ for all t 's, then $\ln(\pi_{it}) = 0$ for all t 's and (2.5) equals (2.4), can consequently, the model used in estimation will correspond to the true conditional choice model.

More succinctly, given true model (2.3), the likelihood function obtained from model (2.4) will mistakenly ignore a sequence of (i, t) -specific fixed effects if and only if at least one choice set $S_{it} = s_t$ of the sequence $S_i = s$ includes at least one alternative not originally in $CS_{it}^* = c_t$. The $\ln(\pi_{it}(\theta))$'s are (i, t) -specific fixed effects that cause inconsistency in estimation. $\pi_{it}(\theta)$ measures the probability that individual i would choose one of the alternatives in her true choice set when faced with the set of products assumed by the researcher. If $s_t \cap c_t \subset s_t$, i.e. if s_t includes alternatives not in c_t , this probability will be strictly less than one, and smaller (and thus the likely inconsistency greater) the more likely it is that i would have preferred one of the products mistakenly included in s_t .

2.3. Quantifying the size of the bias in MNL model

Table 1 provides Monte Carlo evidence on the size of the bias in MNL and mixed MNL models from mistakenly attributing to individuals alternatives that were not available to them. The three panels describe the relative importance of different features of the choice set generating process on the extent of the bias arising from unobserved choice set heterogeneity. We report the average bias and the standard deviation of the estimates (across 20 replications) that arise if the researcher imputes the full choice set of five alternatives instead of the true (heterogeneous and unobserved) choice set to all individuals in all choice situations.

In each scenario, the data generating process is a MNL model with systematic utility $V(X_{ijt}, \theta) = X_{ijt}\beta$ and heterogeneous choice sets across individuals. Given these data, we report in the first column estimates from a MNL model with systematic utility $V(X_{ijt}, \theta) = \delta_{jt} + X_{ijt}\beta$ and with full choice sets, and in the second column estimates from a mixed MNL model with $V(X_{ijt}, \theta_i) = \delta_{jt} + X_{ijt}\beta_i$, $\beta_i = \beta + \sigma \times v_i$, v_i distributed standard normal, and with full choice sets.⁹ Note that, differently from the data generating process, both estimated models include alternative-specific constants δ_{jt} 's, but that this does not yield consistent estimates. In the top panel, we report results for the baseline model, where all individuals make choices from the full choice set. In this case, both the MNL and the mixed MNL models with full choice sets are correctly specified and virtually unbiased. In the second panel, an increasing share of individuals make choices from a choice set of four randomly selected alternatives. In both the standard and mixed MNL models, the bias increases with the share of individuals facing constrained choice sets. In the mixed MNL model the bias in both the estimated mean $\hat{\beta}$ and the estimated standard deviation $\hat{\sigma}$ of the random coefficient increases. In the third panel, 30% of individuals make choices from a choice set of two, three, or four randomly selected alternatives. For a given share of individuals with constrained choice sets, in both estimated models the bias increases with the severity of the constraint in choice sets. In the bottom panel, 10% of individuals have their first-best alternative removed from the choice set, with an increasing "distance" between the systematic utilities of the (removed) first best and the (chosen) second best. The bias increases the more individuals prefer the alternatives that are not included in their true choice sets but that are mistakenly included in the choice sets of the models used in estimation. Interestingly, and differently from the previous two panels, the bias in the mixed MNL estimates is larger in the estimated standard deviation $\hat{\sigma}$ of the random coefficient, while the mean $\hat{\beta}$ is basically unbiased.

The results are intuitive and show the consequences of failing to account for unobserved choice set heterogeneity. The size of the bias can be substantial. Collectively, these results confirm the theoretical insights from Section 2.2: the estimation bias is proportional to the extent of the (incorrect) choice set enlargement and to the size of the systematic utilities of the alternatives mistakenly added to the choice sets. Furthermore, the results show that simply adding alternative-specific constants and random coefficients to a MNL model does not address the econometric problems introduced by unobserved choice set heterogeneity.

3. Two solutions: Integrating over and differencing out heterogeneous unobserved choice sets

The methods proposed in the literature to address unobserved choice set heterogeneity can be grouped into two families: those that "integrate over" and those that "difference out" unobserved choice sets. The first is the approach of Manski (1977), which models the unconditional probability i chooses j by integrating over *all* possible unobserved choice sets that include j . This is akin to treating unobserved choice set heterogeneity in a manner analogous to unobserved preference heterogeneity, and has been widely used in the applied literature (e.g., Goeree, 2008 and Van Nierop et al., 2010).

A second approach that is less commonly used aims at "differencing out" unobserved choice sets by conditioning choice probabilities on subsets of the true choice sets. This approach builds on well known estimators originally developed for other purposes, such as McFadden (1978), Fox (2007), Fox et al. (2011), Hey and Orme (1994), and Dubois et al. (2020) to "difference out" unobserved choice sets in various models (beyond the MNL). We introduce the term "sufficient set" to describe a method of identifying subsets of true choice sets and illustrate how it can be used in models in the "differencing out" approach with increasingly general error structures. We also discuss how the two approaches can be combined to reduce the computational burden of the "integrating over" approach.

3.1. "Integrating over" unobserved choice sets

The most popular approach used in the literature to address unobserved choice set heterogeneity is to jointly model choice set formation and the purchase decision given a choice set. As originally discussed by Manski (1977), the unconditional probability of i selecting choice sequence j can be written as:

$$\Pr[Y_i = j | \theta, \gamma] = \sum_{c \in C_i^*} \Pr[Y_i = j | CS_i^* = c, \theta] \Pr[CS_i^* = c | \gamma], \quad (3.1)$$

⁹ The mixed MNL is estimated by simulated maximum likelihood using 150 shifted and shuffled Halton standard normal draws of v_i per individual. Substantially increasing the number of draws per individual yields results that are qualitatively unchanged.

Table 1
Size of the bias in the MNL model with full choice sets.

	MNL	Mixed MNL	
	Bias (StdDev) $\hat{\beta}$	Bias (StdDev) $\hat{\beta}$	Bias (StdDev) $\hat{\sigma}$
Baseline			
100% of consumers have full choice set	0.005 (0.032)	0.013 (0.034)	0.071 (0.080)
Increasing the share of individuals with constrained choice sets			
90% have full choice set, 10% choose from 4 out of 5	−0.223 (0.021)	−0.019 (0.034)	0.498 (0.023)
70% have full choice set, 30% choose from 4 out of 5	−0.525 (0.013)	−0.272 (0.023)	0.531 (0.019)
50% have full choice set, 50% choose from 4 out of 5	−0.719 (0.007)	−0.529 (0.015)	0.446 (0.015)
Increasing the num. of alt. randomly removed from choice sets			
30% have 4 out of 5 available	−0.525 (0.013)	−0.272 (0.023)	0.531 (0.019)
30% have 3 out of 5 available	−0.839 (0.007)	−0.425 (0.021)	0.683 (0.019)
30% have 2 out of 5 available	−1.139 (0.003)	−0.474 (0.022)	0.916 (0.020)
Increasing the differentiation of alt., 10% have first-best removed			
First-best alternative is slightly better ($V_1 - V_2$)/ $V_1 \simeq 10\%$, $\sigma_X^2 = 1.5$	−0.346 (0.017)	−0.064 (0.031)	0.682 (0.024)
First-best alternative is better ($V_1 - V_2$)/ $V_1 \simeq 35\%$, $\sigma_X^2 = 2.5$	−0.471 (0.012)	−0.026 (0.029)	0.775 (0.020)
First-best alternative is much better ($V_1 - V_2$)/ $V_1 \simeq 110\%$, $\sigma_X^2 = 5.5$	−0.740 (0.009)	0.010 (0.036)	0.891 (0.025)

We consider a population of 1000 individuals making a sequence of choices over 10 choice situations. On each choice situation they choose between a maximum of five alternatives. The indirect utility of each alternative is specified as in Eq. (2.1). The true systematic utility is $V(X_{ijt}, \theta) = X_{ijt}\beta$, and the unobserved portion of utility, ϵ_{ijt} , is distributed i.i.d. Gumbel. In the baseline specification, X_{ijt} is drawn from a normal distribution with mean 0 and variance 5, and $\beta = 2$. In the MNL column, we report estimates of a MNL model with $V(X_{ijt}, \theta) = \delta_{it} + X_{ijt}\beta$ and where all individuals are incorrectly assumed to always have full choice sets. In the mixed MNL column, we report estimates of a mixed MNL model with $V(X_{ijt}, \theta_i) = \delta_{it} + X_{ijt}\beta_i$, $\beta_i = \beta + \sigma \times v_i$, where we draw 150 shifted and shuffled Halton v_i 's per individual from a standard normal, and all individuals are incorrectly assumed to always have full choice sets. The mixed MNL is exclusively used in estimation, the data are always generated from a standard MNL model. The table reports averages of the biases and standard deviations of the estimates across 20 replications per scenario. In the top panel, all individuals make choices from the full choice set. In the second panel, an increasing share of individuals make choices from a choice set of four randomly selected alternatives. In the third panel, 30% of individuals make choices from a choice set of two, three, or four randomly selected alternatives. In the bottom panel, 10% of individuals have their first-best alternative removed from the choice set, with an increasing "distance" between the systematic utilities of the (removed) first best and the (chosen) second best. In the second to bottom panels, choice sets differ across individuals but are constant across choice situations within individual.

where C_i^* is the collection of possible sets of choice sequences to which individual i can be matched. By having information on the matching process between individuals and choice sets, one can integrate over unobserved choice set heterogeneity in a manner analogous to how it is routinely done with unobserved preference heterogeneity. Until recently, it was believed that the identification of model (3.1) relied on the availability of auxiliary data about $\Pr[CS_i^* = c | \gamma]$ (e.g., Roberts and Lattin, 1991) and/or the availability of instruments that exclusively affected the matching between individuals and choice sets (e.g., Goeree, 2008). In a recent paper, however, Abaluck and Adams (2017) present identification results for this model that do not require the availability of such auxiliary data.¹⁰

Whenever model (3.1) is correctly specified and C_i^* is not too large, so that estimation is computationally feasible (e.g., by a Simulated Maximum Likelihood Estimator of (3.1)), then one can use this approach to learn about both preferences θ and about the matching process between individuals and choice sets γ . Knowledge of both θ and γ is essential in many contexts, especially when the researcher is interested in simulating counterfactuals that may involve the re-matching of choice sets to individuals (e.g., Gaynor et al., 2016).

Even though Manski (1977)'s approach represents the best option in many instances, there are also cases in which it may not be appropriate. The rest of this subsection discusses three of the main possible drawbacks of the integrating over approach. The first is that in addition to $\Pr[Y_i = j | CS_i^* = c, \theta]$, the practical implementation of model (3.1) requires knowledge also of the functional form of $\Pr[CS_i^* = c | \gamma]$. The second, perhaps less obvious, is that the collection of choice sets over which expectations are taken is rarely observable and therefore C_i^* may easily be misspecified. Third, even when

¹⁰ The intuition of their argument is that whenever some alternatives are not in the choice sets of some individuals, the discrete-choice analogue of Slutsky symmetry will be violated. They show that one can therefore use deviations from Slutsky symmetry to separately identify γ and θ .

$\Pr[Y_i = j|\theta, \gamma]$ is correctly specified, in practice the estimation of model (3.1) may prove difficult. Indeed, this model suffers from a curse of dimensionality related to the number of elements in C_i^* , which grows exponentially in the number of alternatives available to individual i , J_i , for example, [Abaluck and Adams \(2017\)](#) limit their estimations to the case of $J_i=10$. Two possible ways of alleviating the practical consequences of this curse of dimensionality are the importance sampling procedure proposed by [Goeree \(2008\)](#) (detailed in [Appendix A](#)) and the idea proposed in [Section 3.3](#). Note that both simplifications are not “free” and require additional assumptions on the choice set formation process $\Pr[CS_i^* = c|\gamma]$.

We illustrate these potential drawbacks in the context of [Goeree \(2008\)](#). In our notation, [Goeree \(2008\)](#)’s choice model can be written as:

$$\Pr[Y_{it} = j_t|\theta, \gamma] = \sum_{c_t \in C_t^j} \frac{\exp(V(X_{ijt}, \theta))}{\sum_{r_t \in c_t} \exp(V(X_{irt}, \theta))} \left[\prod_{l_t \in c_t} \phi_{ilt}(\gamma) \prod_{k_t \notin c_t} (1 - \phi_{ikt}(\gamma)) \right], \quad (3.2)$$

where C_t^j is the collection of *all* period t choice sets that include product j_t . This relies on:

- [Assumption 1](#), with $\Pr[Y_{it} = j_t | CS_{it}^* = c_t, \theta] = \frac{\exp(V(X_{ijt}, \theta))}{\sum_{r_t \in c_t} \exp(V(X_{irt}, \theta))}$, and
- the additional assumption that consideration of each product is *independent* of the consideration of the other products: $\Pr[CS_{it}^* = c_t | \gamma] = \prod_{l_t \in c_t} \phi_{ilt}(\gamma) \prod_{k_t \notin c_t} (1 - \phi_{ikt}(\gamma))$.

Even given this second assumption, in [Goeree \(2008\)](#) the total number of products is still very large (> 2100), so the non-parametric estimation of all the ϕ ’s is not feasible. To reduce the dimensionality of estimating the matching of consumers to choice sets, she further assumes that:

$$\phi_{ilt}(\gamma) = \frac{\exp(W_{ilt}\gamma)}{1 + \exp(W_{ilt}\gamma)}, \quad (3.3)$$

with W_{ilt} a vector of observable characteristics and γ a vector of parameters measuring the sensitivity of product consideration with respect to these observables.

This implies that every $c_t \in C_t^j$ will have a strictly positive probability in the distribution of choice sets for each (i, t) , so that $\Pr[CS_{it}^* = c_t | \gamma] > 0$ for every (i, t) . However, it may be that for some (i, t) combination $\Pr[CS_{it}^* = c_t | \gamma] = 0$ for some c_t , i.e. c_t is not in the support of the choice set distribution to which individual i can be matched to in period t :

$$\Pr[CS_{it}^* = c_t | \gamma] = \begin{cases} \prod_{l_t \in c_t} \phi_{ilt}(\gamma) \prod_{k_t \notin c_t} (1 - \phi_{ikt}(\gamma)) & \text{if } c_t \in C_{it}^{j*} \\ 0 & \text{if } c_t \in C_t^j \setminus C_{it}^{j*}, \end{cases} \quad (3.4)$$

where C_{it}^{j*} is the collection of choice sets to which individual i can possibly be matched to in period t . Since C_{it}^{j*} is typically unobserved and heterogeneous across (i, t) combinations, model (3.2) and (3.3) will suffer from support misspecification whenever there exist observations where the true collection of unobserved choice sets to which an individual can be matched to is restricted, i.e. $\exists (i, j, t)$ combination such that $C_{it}^{j*} \subset C_t^j$. With support misspecification, standard estimators will be inconsistent.

To be clear, computing expectations over the power set of the universal set, as with C_t^j in (3.2), would *not* be a problem if one could afford to estimate a truly flexible specification for ϕ that was able to accommodate $\Pr[CS_{it}^* = c_t | \gamma] = 0$ whenever necessary. The problem arises because one is not usually able to estimate a truly flexible model for ϕ , and needs to make additional assumptions along the lines of (3.3). Taken together, C_t^j in (3.2) and (3.3) may introduce bias due to the potential inclusion of infeasible choice sets.¹¹

3.2. “Differencing out” unobserved choice sets

When [Manski \(1977\)](#)’s approach is not appropriate, one can still hope to estimate the preference parameters θ by *differencing out* unobserved choice sets. We start by introducing the differencing out method in the context of the MNL model, then in the next two subsections we extend the discussion to more general models.

¹¹ Two other papers that take a similar approach to [Goeree \(2008\)](#) are [Van Nierop et al. \(2010\)](#) and [Draganska and Klapper \(2011\)](#). [Van Nierop et al. \(2010\)](#) assume that the model for $\Pr[CS_{it}^* = c_t | \gamma]$ is a J -dimensional multivariate normal distribution, which again cannot be 0 for any c_t . In the corresponding equation to (3.2), they compute expectations over C_t^j , associating positive mass to every $c_t \in C_t^j$ rather than only to each $c_t \in C_{it}^{j*} \subset C_t^j$. [Draganska and Klapper \(2011\)](#) assume $\Pr[CS_{it}^* = c_t | \gamma]$ to be a MNL model with choice set C_t^j , and compute expectations over C_t^j , where again the multinomial logit model cannot be exactly 0 for any $c_t \in C_t^j$.

Table 2
Performance of sufficient set logits.

	MNL, full % Bias	MNL, true % Bias	FPH SSL % Bias	PPH SSL % Bias	CP SSL % Bias
Baseline					
100% full choice set	0.3%	0.3%	0.6%	0.8%	1.2%
Increasing share of individuals with a random product removed from choice set					
10% constrained	11.2%	0.2%	0.4%	0.9%	0.8%
30% constrained	26.3%	0.3%	0.6%	1.2%	0.6%
50% constrained	36.0%	0.4%	0.7%	1.0%	0.9%
Increasing share of products randomly removed from choice set					
30% have 4 of 5	26.3%	0.3%	0.6%	1.2%	0.6%
30% have 3 of 5	36.0%	0.7%	1.0%	1.3%	1.2%
30% have 2 of 5	57.0%	0.4%	0.5%	0.7%	0.5%

We consider a population of 1000 individuals making a sequence of choices over 10 choice situations. On each choice situation, they choose between a maximum of five alternatives. The indirect utility of each alternative is specified as in Eq. (2.1). The systematic utility is $V(X_{ijt}, \theta) = \delta_{jt} + X_{ijt}\beta$, and the unobserved portion of utility, ϵ_{ijt} , is distributed i.i.d. Gumbel. X_{ijt} is drawn from a normal distribution with mean 0 and variance 5, $\delta_{jt} = 0$ for all j_t 's, and $\beta = 2$. The table reports averages of the percentage absolute bias of the estimates, $|(\hat{\beta} - \beta)/\beta| \times 100$. In the top panel, all individuals make choices from the full choice set. In the central panel, an increasing share of individuals make choices from a choice set of four randomly selected alternatives. In the bottom panel, 30% of individuals make choices from a choice set of two, three, or four randomly selected alternatives. In the central and bottom panels, choice sets differ across individuals but are constant across choice situations within individual. We simulate and average results over 20 replications per scenario. To speed up computations, the CP SSL is estimated by sampling at random (uniformly), for each individual, 5000 permutations of the observed sequence of choices, as suggested by D'Haultfoeulle and Iaria (2016).

3.2.1. Multinomial logit model

This approach originates from McFadden (1978)'s idea of estimating MNL models from subsets of the observed but potentially large choice sets. To implement this idea when choice sets are unobserved, inspired by Chamberlain (1980), we combine information on i 's observed choice sequence in panel data environments, $Y_i = j$, to construct subsets of i 's true but unobserved choice set, $\mathcal{CS}_i^* = c$, and then rely on McFadden (1978)'s consistent estimator of θ from subsets of the true choice sets. We call these subsets *sufficient sets*: they are a collection of choice sequences generated by any correspondence $f(Y_i)$ that satisfies the following property.

Condition 1. Given any choice sequence $Y_i \in \mathcal{CS}_i^*$, the correspondence f is such that $Y_i \in f(Y_i)$ and $f(Y_i) \subseteq \mathcal{CS}_i^*$.

It is easy to see that if Assumption 1 (i.e., MNL model) and Condition 1 hold, then $f(Y_i)$ will be a sufficient statistic for \mathcal{CS}_i^* or, equivalently, a MNL model conditional on $f(Y_i)$ will be guaranteed to satisfy the IIA property even if choice sets are unobserved. It is for this reason that we call any f that satisfies Condition 1 a sufficient set. More precisely, if Assumption 1 and Condition 1 hold, then for every individual i and choice sequence $Y_i = j$ such that $f(j) = r$:

$$\Pr[Y_i = j | f(Y_i) = r, \theta] = \frac{\prod_{t=1}^T \exp(V(X_{ijt}, \theta))}{\sum_{k \in f(Y_i)=r} \prod_{t=1}^T \exp(V(X_{ikt}, \theta))} \quad (3.5)$$

and θ can be consistently estimated by the conditional Maximum Likelihood Estimator derived from $\Pr[Y_i = j | f(Y_i) = r, \theta]$; see Appendix B for details. Eq. (3.5) is a direct consequence of the IIA property. We define the MNL in (3.5) as the **Sufficient Set Logit (SSL)** model.

In essence, one can estimate preferences θ based only on the variation in characteristics of those products in i 's sufficient set, rather than on her full (but unobserved) sequence of choice sets. This is evident since Eq. (3.5) does not depend on i 's unobserved sequence of choice sets, $\mathcal{CS}_i^* = c$. Whenever $\mathcal{CS}_i^* = c$ is observed, the econometrician can easily detect appropriate subsets of $\mathcal{CS}_i^* = c_t$ for any i in t , and rely on McFadden (1978) to consistently estimate θ . However, as we saw in Section 2.2, whenever \mathcal{CS}_i^* is unobserved the econometrician needs to be careful in constructing choice sets for each i in any t so that they are *actual subsets* of each \mathcal{CS}_i^* . The main idea of the differencing out approach is to exploit individual i 's observed choice sequence Y_i , paired with assumptions about the evolution of \mathcal{CS}_i^* over t , to construct a proper subset of \mathcal{CS}_i^* , and then to rely on McFadden (1978) to consistently estimate θ .

In Table 2, we present Monte Carlo evidence on the effectiveness of the SSL for estimating preferences in the presence of unobserved choice set heterogeneity. The rows in Table 2 characterize different types of unobservable choice set heterogeneity and the columns show the % bias arising from alternative estimators: the MNL on the full choice set, the MNL on the true choice set, and alternative versions of the SSL, with different sufficient sets corresponding to those introduced in Section 4.1. Going down the first column indicates that increasing unobserved choice set heterogeneity, either in the form of increasing the share of individuals with a product missing from their specified choice set or in the form of an increasing number of products missing from their specified choice set, is associated with significant bias that is not present either when choice set heterogeneity is correctly specified (in column 2), or when using sufficient sets

that are subsets of consumers' true choice sets (columns 3–5). Further details of this Monte Carlo exercise, as well as an extension to two economically relevant choice set formation processes – a model of consumer screening and a model of costly search – are provided in [Appendix D](#).

Discussion. The differencing out approach is quite general and works for any f that generates subsets of \mathcal{CS}_i^* . In [Section 4.1](#) we discuss how different economic models can be used to motivate alternative specifications of sufficient sets. While these examples of sufficient sets are suggestive, we note here that they represent a set of sufficient conditions that imply the SSL in [\(3.5\)](#), but they are neither necessary nor the minimal sufficient conditions for the result to hold. In [Section 4.3.1](#) we discuss statistical tests to help researchers choose among different sufficient sets.

The SSL expression in [\(3.5\)](#) makes clear that individual i will have a non-zero log-likelihood contribution whenever her observed choice sequence $Y_i = j$ gives rise to a non-singleton $f(j) = r$. In the practical examples of sufficient sets we discuss in the paper, similar to [Chamberlain \(1980\)](#), this happens whenever the observed choice sequence $Y_i = (j_1, \dots, j_t, \dots, j_T)$ entails “some” switching, so that there exist at least two elements j_t and $j_{t'}$ in the sequence for which $j_t \neq j_{t'}$. All those observations for which only one alternative is chosen repeatedly throughout the sequence will be dropped from the log-likelihood function.

Note that, in practice, the SSL is a MNL in which the choice set is given by a (potentially huge) set of choice sequences over T choice situations, $f(Y_i) = r$. While this can be numerically inconvenient for some peculiar $f(Y_i)$, it will usually be possible to re-express [\(3.5\)](#) in a way that greatly simplifies its practical implementation. In particular, the SSL in [\(3.5\)](#) over choice sequences can be equivalently expressed as the product of T separate t -specific MNL's over alternatives if and only if the sufficient set (over choice sequences) $f(Y_i)$ can be expressed as the cartesian product of T separate t -specific sufficient sets (over alternatives). In other words:

$$\begin{aligned} \Pr[Y_i = j | f(Y_i) = r, \theta] &= \prod_{t=1}^T \Pr[Y_{it} = j_t | f_t(Y_i) = r_t, \theta] \\ &= \prod_{t=1}^T \frac{\exp(V(X_{ijt}, \theta))}{\sum_{k_t \in f_t(k)=r_t} \exp(V(X_{ikt}, \theta))}, \end{aligned} \quad (3.6)$$

if and only if $f(Y_i) = \times_{t=1}^T f_t(Y_i)$; for a derivation see [Appendix C](#). To see why expression [\(3.6\)](#) results in more convenient estimators than expression [\(3.5\)](#), suppose that the econometrician specifies a $f(Y_i)$ such that in each $t = 1, \dots, 10$ an individual can choose one out of $J_t = J = 5$ different alternatives. It follows that $f(Y_i)$ contains 5^{10} possible choice sequences of length $T = 10$. By using the SSL in [\(3.5\)](#), the econometrician would have to estimate a huge MNL model with a summation over 5^{10} addends in the denominator. However, since $f(Y_i)$ can be obtained as the cartesian product of $T = 10$ separate t -specific sets each containing $J = 5$ alternatives, expression [\(3.6\)](#) guarantees that this SSL model can equivalently be expressed as the product of $T = 10$ MNL models each with a summation over $J = 5$ addends in the denominator. The examples of sufficient sets that we propose in [Section 4.1](#) all satisfy this condition, giving rise to computationally simple estimators. As we will illustrate later, a famous exception to the equivalence between [\(3.5\)](#) and [\(3.6\)](#) is the Choice Permutations sufficient set first proposed by [Chamberlain \(1980\)](#), which defines – in our terminology – the sufficient set as the collection of all permutations of the observed $Y_i = j$ and which results in [Chamberlain \(1980\)](#)'s fixed effect logit model.¹²

3.2.2. Individual-specific logit and mixed logit models

In this subsection, we illustrate how sufficient sets can be used in the context of MNL models with individual-specific preference parameters. We start by discussing the simplest case in which the econometrician observes a large T for each individual. We then move on to the more complex scenario in which the econometrician only observes a small T for each individual: after a brief introduction of the model and basic challenges, we focus on a discrete mixture version of it for which both identification and estimation can be discussed in simple and intuitive terms.

Individual-specific MNL model: panel data with large T

When the econometrician observes a large number of choice situations per individual, $T \rightarrow \infty$ for fixed I , then she can rely on the results from the previous subsection and on the estimator proposed by [Hey and Orme \(1994\)](#) and [Dubois et al. \(2020\)](#) to estimate a separate MNL model with unobserved choice sets for each individual.

Assumption 2(a). Suppose that each i has systematic utilities of the form $V(X_{ijt}, \theta_i)$ with individual-specific preferences θ_i distributed according to $p(\theta_i = \theta | \psi)$, where ψ is a vector of parameters. Conditional on all $V(X_{ijt}, \theta_i)$'s, on $\theta_i = \theta$, and on $\mathcal{CS}_i^* = c$, ϵ_{ijt} from [\(2.1\)](#) is distributed i.i.d. Gumbel.

¹² In those cases in which $f(Y_i)$ cannot be expressed as $\times_{t=1}^T f_t(Y_i)$, the econometrician can directly apply [McFadden \(1978\)](#) and estimate θ from subsets of the observed but potentially huge sufficient sets $f(Y_i)$. As an example, [D'Haultfoeille and Iaria \(2016\)](#) implement this idea in the context of [Chamberlain \(1980\)](#)'s fixed effect logit model. Matlab codes are available on the authors' personal webpages.

Irrespective of the distribution of θ_i , when $f(Y_i) = \times_{t=1}^T f_t(Y_i)$, it follows from the previous subsection that i 's probability of choosing $Y_i = j$ conditional on $f(Y_i) = r$ and on θ_i is given by:

$$\Pr[Y_i = j | f(Y_i) = r, \theta_i] = \prod_{t=1}^T \frac{\exp(V(X_{ijt}, \theta_i))}{\sum_{k \in f_t(k)=r_t} \exp(V(X_{ikt}, \theta_i))}, \quad (3.7)$$

which we call the **Individual Sufficient Set Logit (ISSL)** to distinguish it from the standard SSL from (3.6). Even though the realization of the random coefficients θ_i is unobserved and potentially heterogeneous across individuals, when $T \rightarrow \infty$ the ISSL model (3.7) can be directly used as an individual-specific likelihood function on the basis of which to construct a consistent individual-specific MLE of θ_i . In other words, for any finite I , when T is large the econometrician can treat the unobserved θ_i as a “fixed effect” and estimate it from the individual-specific MLE derived from (3.7), for each $i = 1, \dots, I$. Hey and Orme (1994) and Dubois et al. (2020) propose this idea for situations with observed choice sets, but the results from the previous subsection imply that their methods – when combined with sufficient sets from Condition 1 – can be readily applied also to situations with unobserved choice sets. At an intuitive level, by following this procedure, the econometrician would estimate by MLE a separate θ_i for each individual $i = 1, \dots, I$ and then – if in addition to $T \rightarrow \infty$, also $I \rightarrow \infty$ – she could recover non-parametrically the distribution of random coefficients $p(\theta_i = \theta | \psi)$ by simply computing the frequency of each realization $\theta_i = \theta$ among the estimates. For more details about this estimation procedure, see Dubois et al. (2020).

Discrete distribution of random coefficients: small T

Here we discuss the typical and more complex scenario in which the econometrician only observes a small number of choice situations T per individual, with $I \rightarrow \infty$. In this case the econometrician will instead need to treat θ_i as a random effect and, relying on its distribution $p(\theta_i = \theta | \psi)$, integrate it over to derive i 's choice probability.

Given Assumption 2(a) and Condition 1, by conditioning the probability of choice sequence $Y_i = j$ on the sufficient set $f(Y_i) = r$, with $f(Y_i) = r \subseteq CS_i^* = c$, we obtain the **Sufficient Set Mixed Logit (SSML)**:

$$\begin{aligned} \Pr[Y_i = j | f(Y_i) = r, \psi] \\ &= \int_{\theta} \Pr[Y_i = j, \theta_i = \theta | f(Y_i) = r, \psi] d\theta \\ &= \int_{\theta} \Pr[Y_i = j | f(Y_i) = r, \theta_i = \theta] p(\theta_i = \theta | f(Y_i) = r, \psi) d\theta \\ &= \int_{\theta} \frac{\prod_{t=1}^T \exp(V(X_{ijt}, \theta))}{\sum_{k \in f(Y_i)=r} \prod_{t=1}^T \exp(V(X_{ikt}, \theta))} p(\theta | f(Y_i) = r, \psi) d\theta. \end{aligned} \quad (3.8)$$

Note from (3.8) that the distribution of random coefficients used to integrate over unobserved preference heterogeneity is *conditional* on the realization of the sufficient set $f(Y_i) = r$, i.e. $p(\theta_i = \theta | f(Y_i) = r, \psi)$. As a consequence, and differently from the unconditional mixed logit model commonly estimated by applied researchers, the SSML in (3.8) is a conditional mixed logit model which requires the specification of the conditional distribution of random coefficients $p(\theta_i = \theta | f(Y_i) = r, \psi)$, as opposed to the more standard unconditional distribution of random coefficients $p(\theta_i = \theta | \psi)$. The two are related by:

$$p(\theta_i = \theta | \psi) = \sum_r p(\theta_i = \theta | f(Y_i) = r, \psi) \Pr[f(Y_i) = r], \quad (3.9)$$

where the probability of each realization r of the sufficient set, $\Pr[f(Y_i) = r]$, is observed in the data. Given (3.9), it is apparent that parametric assumptions on the unconditional distribution of random coefficients, $p(\theta_i = \theta | \psi)$, will not typically translate into convenient restrictions on the conditional distributions, $p(\theta_i = \theta | f(Y_i) = r, \psi)$. For example, if $p(\theta_i = \theta | \psi)$ is a normal density, that certainly does not imply that $p(\theta_i = \theta | f(Y_i) = r, \psi)$ will also be a normal density. This complexity is a consequence of the sample selection on the realized random coefficients introduced by the conditioning $f(Y_i) = r$. The next example provides some intuition about this sample selection problem.

Example on sample selection of random coefficients. Consider a market with two kinds of products: high-quality/high-price products, collected in set h , and low-quality/low-price products, collected in set ℓ . Suppose that the *unconditional* distribution of the marginal utility of “quality” β_i in the population is discrete with two points of support, β_{high} with probability $\Pr[\beta_i = \beta_{high} | \theta] = \theta_{high}$ and β_{low} with probability $\Pr[\beta_i = \beta_{low} | \theta] = \theta_{low} = 1 - \theta_{high}$, where $\beta_{high} > \beta_{low}$. The parameters β_{high} , β_{low} , and θ_{high} are the objects of interest. Suppose that all the choice sequences (Y_i) observed in the data give rise to two realizations of the sufficient set: $f(Y_i) = h$, collecting all the high-quality/high-price products, and $f(Y_i) = \ell$, collecting the remaining low-quality/low-price products. Assume that $f(Y_i) = h$ is observed in the data with probability $\Pr[f(Y_i) = h] = p$ and $f(Y_i) = \ell$ with probability $\Pr[f(Y_i) = \ell] = 1 - p$. Among the i 's with $f(Y_i) = h$, the *conditional* distribution of the random coefficients is $\Pr[\beta_i = \beta_{high} | f(Y_i) = h] = \rho_{high}^h$ and $\Pr[\beta_i = \beta_{low} | f(Y_i) = h] = \rho_{low}^h = 1 - \rho_{high}^h$.

Similarly, among the i 's with $f(Y_i) = \ell$, it is $\Pr[\beta_i = \beta_{high} | f(Y_i) = \ell] = \rho_{high}^\ell$ and $\Pr[\beta_i = \beta_{low} | f(Y_i) = \ell] = \rho_{low}^\ell = 1 - \rho_{high}^\ell$. For any i with $f(Y_i) = h$, the conditional choice probability of choosing sequence $Y_i = j \in f(Y_i) = h$ is:

$$\begin{aligned} \Pr[Y_i = j | f(Y_i) = h, \beta_{high}, \beta_{low}, \rho_{high}^h] \\ = \Pr[Y_i = j | f(Y_i) = h, \beta_i = \beta_{high}] \times \rho_{high}^h + \Pr[Y_i = j | f(Y_i) = h, \beta_i = \beta_{low}] \times (1 - \rho_{high}^h), \end{aligned} \quad (3.10)$$

where $\Pr[Y_i = j | f(Y_i) = h, \beta_i = \beta_{high}]$ and $\Pr[Y_i = j | f(Y_i) = h, \beta_i = \beta_{low}]$ are MNL formulas with choice set $f(Y_i) = h$ and evaluated at preference parameters β_{high} and β_{low} , respectively. For any i with $f(Y_i) = \ell$, then $\Pr[Y_i = j | f(Y_i) = \ell, \beta_{high}, \beta_{low}, \rho_{high}^\ell]$ is analogous to (3.10). Eq. (3.9) linking the conditional to the unconditional distribution of random coefficients simplifies to $\theta_{high} = \rho_{high}^h \times p + \rho_{high}^\ell \times (1 - p)$ and $\theta_{low} = 1 - \theta_{high}$. In this example, absence of sample selection corresponds to $\rho_{high}^h = \rho_{high}^\ell = \theta_{high}$. This condition would be violated, for instance, if consumers with a higher preference for quality $\beta_i = \beta_{high}$ were more likely to be observed purchasing high-quality products, so that $\rho_{high}^h > \rho_{high}^\ell$.

This selection problem greatly complicates general treatments about identification and estimation of the SSML model (3.8), see for example Keane and Wasi (2012). However, there are simple versions of (3.8) whose identification can be readily shown and estimation easily performed in practice.

We now turn to one of these cases, where both the random coefficients θ_i and the regressors X_i have a discrete, finite, and known support. In this scenario, the identification of the model can be described in very transparent terms, while the estimation can be performed by Ordinary Least Squares (OLS) or, to improve efficiency, by the easy-to-implement inequality-constrained least square estimator proposed by Bajari et al. (2007) and Fox et al. (2011).

Assumption 2(b). Suppose the matrix $X_i = [X_{i1}, \dots, X_{it}, \dots, X_{iT}]$ gathers all of i 's regressors over the T choice situations, $\Theta = \{\theta_1, \dots, \theta_q, \dots, \theta_Q\}$ is the discrete and finite support of the random coefficients θ_i , and $\Psi^r = [\psi_1^r, \dots, \psi_q^r, \dots, \psi_Q^r]'$ the associated conditional weights or conditional probability masses. There is a finite number P of different values taken by the regressors X_i , so that any $X_i \in \{X_1, \dots, X_p, \dots, X_P\}$. Both Θ and $\{X_1, \dots, X_p, \dots, X_P\}$ are known to the econometrician.

Given the additional Assumption 2(b), the SSML model (3.8) simplifies to:

$$\Pr[Y_i = j | X_i, f(Y_i) = r, \Theta, \Psi^r] = \sum_{q=1}^Q \Pr[Y_i = j | X_i, f(Y_i) = r, \theta_i = \theta_q] \times \psi_q^r. \quad (3.11)$$

For the purpose of identification, the left hand side of (3.11), $\Pr[Y_i = j | X_p, f(Y_i) = r, \Psi^r]$, is known for any combination (j, p, r) . Because $\Theta = \{\theta_1, \dots, \theta_q, \dots, \theta_Q\}$ is known, each $\Pr[Y_i = j | X_p, f(Y_i) = r, \theta_i = \theta_q]$ from (3.11) is also known for any (j, X_p, r, θ_q) combination (i.e., a simple SSL with parameters θ_q). For given r , as a consequence, identification boils down to guaranteeing the existence of a unique solution Ψ^r to the system of linear equations in (3.11).

There are Q sufficient set logit probabilities on the right hand side of (3.11). For brevity, we call each of them $\Pr[j | X_p, r, \theta_q]$ and re-write (3.11) as:

$$\begin{aligned} \Pr[Y_i = j | X_p, f(Y_i) = r, \Theta, \Psi^r] &= \sum_{q=1}^Q \Pr[j | X_p, r, \theta_q] \times \psi_q^r \\ &= \mathbf{Pr}[j | X_p, r, \Theta] \Psi^r, \end{aligned} \quad (3.12)$$

where $\mathbf{Pr}[j | X_p, r, \Theta] = [\Pr[j | X_p, r, \theta_1], \dots, \Pr[j | X_p, r, \theta_Q]]$ and $\Psi^r = [\psi_1^r, \dots, \psi_q^r, \dots, \psi_Q^r]'$. Denote the number of choice sequences in $f(Y_i) = r$ by S_r (i.e., $|r| = S_r$). For each (X_p, r) combination, with $p = 1, \dots, P$, we have a system of S_r equations like (3.12) and, in turn, we have P such systems of S_r equations (one for each X_p). By stacking all of these equations together, we obtain a (potentially huge) system of $P \cdot S_r$ equations for any r :

$$\begin{aligned} \mathbf{Pr}[\mathbf{X}, r, \Theta, \Psi^r] &= \begin{bmatrix} \mathbf{Pr}[X_1, r, \Theta] \\ \vdots \\ \mathbf{Pr}[X_p, r, \Theta] \\ \vdots \\ \mathbf{Pr}[X_P, r, \Theta] \end{bmatrix} \Psi^r \\ &= \mathbf{Pr}[\mathbf{X}, r, \Theta] \Psi^r, \end{aligned} \quad (3.13)$$

where $\mathbf{X} = [X_1, \dots, X_p, \dots, X_P]$, $\mathbf{Pr}[\mathbf{X}, r, \Theta, \Psi^r]$ is the $P \cdot S_r \times 1$ vector that stacks together all the observed $\Pr[Y_i = j | X_p, f(Y_i) = r, \Theta, \Psi^r]$ conditional choice probabilities, for all X_p 's and all $j \in f(Y_i) = r$. For any $p = 1, \dots, P$, $\mathbf{Pr}[X_p, r, \Theta]$ is a $S_r \times Q$ matrix with rows given by $\mathbf{Pr}[j | X_p, r, \Theta]$, for $j = 1, \dots, S_r$. $\mathbf{Pr}[\mathbf{X}, r, \Theta]$ is the $P \cdot S_r \times Q$ matrix that stacks together all the P matrices $\mathbf{Pr}[X_p, r, \Theta]$ with $p = 1, \dots, P$.

It is then immediate to see that given [Assumptions 2\(a\)](#), [2\(b\)](#), and [Condition 1](#), the discrete conditional distribution of random coefficients Ψ^r , conditional on sufficient set $f(Y_i) = r$, is identified whenever $\mathbf{Pr}[\mathbf{X}, r, \Theta]$ is of full column rank:

$$\Psi^r = (\mathbf{Pr}[\mathbf{X}, r, \Theta]' \mathbf{Pr}[\mathbf{X}, r, \Theta])^{-1} \mathbf{Pr}[\mathbf{X}, r, \Theta]' \mathbf{Pr}[\mathbf{X}, r, \Theta, \Psi^r]. \quad (3.14)$$

This full column rank condition is sufficient but not necessary for the identification of Ψ^r .¹³ An obvious necessary condition for $\mathbf{Pr}[\mathbf{X}, r, \Theta]$ to be of full column rank is $Q \leq P \cdot S_r$. In other words, the rank condition embedded in (3.14) requires one to have “enough” measurements in the sense of many potential choice sequences (a high S_r) and large variation in the regressors (a high P). Intuitively, when $P \cdot S_r$ is high, then it is “easier” to sustain a finer grid of points Θ (a high Q). In addition, at the potential cost of some loss of efficiency, Eq. (3.14) can be used in isolation to recover Ψ^r for each or only for some of the realizations of the sufficient set $r = 1, \dots, R$.

In practice, these features suggest to focus on the sub-sample of observations corresponding to realizations of the sufficient set r 's with high $P \cdot S_r$, so to be able to identify and estimate a fine grid Θ with a high Q . This is important because, given any Q , for all those r 's with $Q > P \cdot S_r$, the vector of weights Ψ^r may not be identified (i.e., $\mathbf{Pr}[\mathbf{X}, r, \Theta]$ will not be of full column rank). This highlights a trade-off in the choice of the number of grid points Q . On the one hand, the larger the Q the more credible the model, to the extreme of being able to approximate even continuous mixing distributions (see [Fox et al., 2016](#)), but at the cost of having to use potentially only a small part of the full sample, i.e. those individuals with realizations of $f(Y_i)$ with large $P \cdot S_r$. On the other hand, with a small Q one may be able to use a larger portion of the sample, but the risk of misspecification will be higher. This trade-off is salient because only by having an estimate of Ψ^r for all $r = 1, \dots, R$ one can recover the *unconditional* distribution of random coefficients, the object typically estimated in standard mixed logit models. By allowing for this further dimension of unobserved heterogeneity without any additional data, one will be able to learn less about the distribution of random coefficients.

Eq. (3.14) readily leads to a simple estimator: a separate (and potentially huge) OLS estimator for each $r = 1, \dots, R$. Each individual OLS will provide an estimate of the vector of weights Ψ^r (the distribution of random coefficients conditional on $f(Y_i) = r$). In the context of the unconditional mixed logit (with known choice sets), such a simple-to-implement estimator was first proposed by [Bajari et al. \(2007\)](#) and further extended by [Fox et al. \(2011\)](#). To improve efficiency, these papers propose an inequality-constrained least square estimator that complements the OLS with the natural constraints implied by Ψ^r being a vector of probabilities.¹⁴ A refined version of this estimator that does not require perfect ex-ante knowledge of the grid Θ was proposed by [Fox et al. \(2016\)](#).

3.2.3. Beyond Gumbel errors: [Fox \(2007\)](#)'s pairwise maximum score estimator

[McFadden \(1978\)](#) showed that it is possible to consistently estimate preferences by a conditional Maximum Likelihood Estimator (MLE) using subsets of individuals' true choice sets, but this was only for the MNL model. More recently, [Bierlaire et al. \(2008\)](#) extended the result to discrete-choice models with block-diagonal Generalized Extreme Value errors. To the best of our knowledge, in the context of cross-sectional or “short” panel data, results of this kind are not available for the nested logit and for the mixed logit models, even though some interesting approximations have been proposed by [Keane and Wasi \(2012\)](#) and [Guevara and Ben-Akiva \(2013a,b\)](#).¹⁵

Building on [Manski \(1975\)](#), [Fox \(2007\)](#) extended ([McFadden, 1978](#)) by showing that semi-parametric discrete-choice models can be consistently estimated with a **Pairwise Maximum Score Estimator (PMSE)** using subsets of individuals' true choice sets. In this subsection we discuss the use of sufficient sets in the context of the PMSE proposed by [Fox \(2007\)](#).

Assumption 3. Suppose that $V(X_{ikt}, \theta) = X_{ikt}\theta$ and that X_{it} is the matrix stacking all the X_{ikt} 's for $k \in CS_{it}^* = c_{it}$. For any given (i, t) and $k, k' \in CS_{it}^* = c_{it}$, $X_{ikt}\theta > X_{ikt'}\theta$ if and only if $\Pr[Y_{it} = k | X_{it}, CS_{it}^* = c_{it}, \theta] > \Pr[Y_{it} = k' | X_{it}, CS_{it}^* = c_{it}, \theta]$.

[Assumption 3](#) ([Fox, 2007](#)'s Assumption 1) states that the alternatives belonging to (i, t) 's true but unobserved choice set with higher systematic utilities are more likely to be chosen. Note that, differently from parametric models such as the MNL or the probit, [Assumption 3](#) does not impose that the distribution of ϵ_{ikt} is the same across individuals or even that the distribution is the same across the choice situations of the same individual (e.g., ϵ_{ikt} could be distributed Laplace while $\epsilon_{ikt'}$ normal). [Goeree et al. \(2005\)](#) show that a sufficient condition for [Assumption 3](#) is that, for any (i, t) , the joint density of the errors across alternatives is *exchangeable*.¹⁶ The implementation of the PMSE with unobserved and heterogeneous choice sets requires an additional condition on sufficient sets.

¹³ There are at least two reasons why the full column rank condition behind (3.14) may be stronger than necessary for identification (i.e., it is possible to achieve identification with $\mathbf{Pr}[\mathbf{X}, r, \Theta]$ of rank less than Q): the possibility of sparsity in Ψ^r (i.e., $\varphi_q^r = 0$ for some q) and the fact that the correspondence f is the same across different realizations $r = 1, \dots, R$ (i.e., the *marginal* distribution of the random coefficients (3.9) imposes restrictions across the R realizations of the sufficient set). We leave the investigation of the necessary conditions for the identification of Ψ^r to future work.

¹⁴ So that for each r : $0 \leq \psi_q^r \leq 1$, $q = 1, \dots, Q$ and $\sum_{q=1}^Q \psi_q^r = 1$.

¹⁵ The lack of general extensions of [McFadden \(1978\)](#) to mixed logit models motivates our focus in the last subsection on individual-specific MNL models in the context of “long” panel data and on mixed logit models with *discrete* distributions of random coefficients when only “short” panel data are available.

¹⁶ Despite the flexibility, there are popular models among applied researchers that violate [Assumption 3](#), such as the mixed logit model. See [Fox \(2007\)](#) for more details about [Assumption 3](#) in general.

Condition 2. Suppose that $f(Y_i) = \times_{t=1}^T f_t(Y_i) = \times_{t=1}^T f_{it}$ and that there is a non-empty set N of (i, t) 's with $|N| = n \leq I \cdot T$ for which $K = \cap_{(i,t) \in N} f_{it}$ contains at least two alternatives, $|K| \geq 2$.

Condition 2 imposes two restrictions. First, it requires the sufficient set over choice sequences $f(Y_i)$ to be the cartesian product of t -specific sufficient sets f_{it} 's over alternatives. Second, it requires that there is a set of (i, t) 's whose sufficient sets f_{it} 's contain the same two or more alternatives. The PMSE makes pairwise comparisons of alternatives belonging to some subset K for all those (i, t) 's that are known to have originally made choices from some choice set CS_{it}^* such that $K \subseteq CS_{it}^*$. **Condition 2** uses sufficient sets to construct a K guaranteed to be strictly included in the true but unobserved choice set CS_{it}^* of every (i, t) belonging to N .¹⁷

With a slight abuse of notation, we re-label the alternatives in subset K so that $K = \{1, \dots, k, \dots, K\}$. The PMSE using choice-based data on the subset K of alternatives is the parameter vector $\hat{\theta}_n^K$ that maximizes:

$$Q_n^K(\theta) = \sum_{k=1}^{K-1} \sum_{k'=k+1}^K \frac{1}{n} \sum_{(i,t) \in N} (1[Y_{it} = k] \cdot 1[X_{ikt}\theta > X_{ik't}\theta] + 1[Y_{it} = k'] \cdot 1[X_{ik't}\theta > X_{ikt}\theta]). \quad (3.15)$$

Given **Assumption 3**, **Conditions 1, 2**, and some additional technical assumptions (i.e. Assumptions 3 and 4 from [Fox, 2007](#) at p.1009 and p.1011), [Fox \(2007\)](#)'s Theorem 1 guarantees that the Pairwise Maximum Score Estimator $\hat{\theta}_n^K$ is consistent for θ .

3.3. Combining sufficient sets with the “integrating over” approach

In Section 3.2, we discussed the use of sufficient sets to specify conditional discrete-choice models that “difference out” unobserved choice sets, but they can also be used to simplify the practical estimation of *unconditional* discrete-choice models that “integrate over” unobserved choice sets, alleviating the curse of dimensionality embedded in [Manski \(1977\)](#)'s approach. An early application of this idea can be found in [Chiang et al. \(1998\)](#).

As discussed in Section 3.1, even when model (3.1) is correctly specified and identified, its estimation is likely to suffer from a curse of dimensionality because the number of elements in C_i^* grows exponentially in the number of alternatives J_i available to individual i (i.e., $|C_i^*| = 2^{J_i} - 1$). A direct consequence of this curse of dimensionality is that, unless the researcher makes strong functional form assumptions on $\Pr[CS_i^* = c | \gamma]$, the model can be estimated only when J_i is small. Sufficient sets can provide additional restrictions on model (3.1) and make its estimation possible if J_i is large (and more tractable for any given J_i). If **Condition 1** is satisfied, $f(Y_i) = r \subseteq CS_i^* = c$, where c is the true set of choice sequences to which i is matched. It therefore follows that any set of choice sequences $c' \in C_i^*$ such that $f(Y_i) = r \not\subseteq CS_i^* = c'$ cannot be the set of choice sequences to which i is matched, so that $\Pr[CS_i^* = c' | \gamma]$ must be zero. In other words, the researcher knows that i 's true but unobserved choice set must contain all the choice sequences in the sufficient set, and consequently any candidate choice set that does not include even just one of these choice sequences can be removed from the collection of possible sets of choice sequences C_i^* .

Let $C_{f(Y_i)} = \{c | f(Y_i) = r \subseteq CS_i^* = c\}$ be the collection of choice sets consistent with $f(Y_i) = r$ (i.e., that include r). Then model (3.1) simplifies to:

$$\Pr[Y_i = j | \theta, \gamma] = \sum_{c \in C_{f(Y_i)}} \Pr[Y_i = j | CS_i^* = c, \theta] \times \Pr[CS_i^* = c | \gamma]. \quad (3.16)$$

where the only difference with (3.1) is in the terms included in the summation. Note that $C_{f(Y_i)}$ will typically be substantially smaller than the unrestricted C_i^* . For example, suppose that there are four possible choice sequences: a , b , c , and d . Depending on their observed choice sequence Y_i , individual i will have a sufficient set of one of four possible sizes: $|f(Y_i)| = 1$ (e.g., $f(Y_i) = \{a\}$, $f(Y_i) = \{b\}$, etc.), $|f(Y_i)| = 2$ (e.g., $f(Y_i) = \{a, b\}$, $f(Y_i) = \{b, c\}$, etc.), $|f(Y_i)| = 3$ (e.g., $f(Y_i) = \{a, b, c\}$, $f(Y_i) = \{b, c, d\}$, etc.), or $|f(Y_i)| = 4$ (i.e., $f(Y_i) = \{a, b, c, d\}$). The collection C_i^* , usually specified as the power set of $\{a, b, c, d\}$, will then contain $2^4 - 1 = 15$ possible (non-empty) choice sets. However, $C_{f(Y_i)}$ will only contain: 8 choice sets if $|f(Y_i)| = 1$, 4 choice sets if $|f(Y_i)| = 2$, 2 choice sets if $|f(Y_i)| = 3$, and 1 choice set if $f(Y_i) = \{a, b, c, d\}$. Importantly, note that this use of the sufficient sets is quite general and does not rely on any functional form assumptions made by the researcher in specifying model (3.1).

3.4. Pros and cons of “integrating over” versus “differencing out”

Each of the two main approaches to the problem of unobserved choice set heterogeneity discussed above presents advantages and disadvantages. While “integrating over” requires additional functional form assumptions and data on the choice set formation process and it is computationally more intensive, it enables researchers to learn about both the preference parameters θ and the choice set formation parameters γ . Learning about both θ and γ may be essential in applications in which the key counterfactuals involve re-matching of choice sets to individuals. In contrast, “differencing out” requires less prior knowledge and data on the choice set formation process and is simpler to implement, but it does not allow the estimation of the parameters γ .

¹⁷ As we will see below, the sizes of N and K directly affect the number of pairwise comparisons, i.e. observations, used to construct the PMSE.

Within the differencing out approach, we discussed four models: the Sufficient Set Logit (SSL) model, the Individual Sufficient Set Logit (ISL) model, the Sufficient Set Mixed Logit (SSML) model, and the semi-parametric Pairwise Maximum Score Estimator (PMSE) studied by Fox (2007). Despite its simplicity and pedagogical value, the SSL model may be unattractive in many applications because of the IIA property. How then to choose among the others?

When data on large T are available, one can opt for the ISL model given its practical simplicity (basically, the estimation of I separate MNL models), or avoid Gumbel preference assumptions and separately estimate Fox (2007)'s semi-parametric model at the level of each individual. The necessary requirement of a large T may however prevent the use of the ISL model in some applications. When only small T data are available, the choice is between the SSML and Fox (2007)'s PMSE. The primary advantage of the PMSE is to allow for flexible distributions of unobserved preferences within the boundaries of Assumption 3. By contrast, the SSML model requires the distribution of random coefficients to be discrete and its support Θ to be known in advance (see Assumption 2(b)).

Researchers using any of the models discussed here can evaluate the willingness to pay for alternatives' attributes (see, for example, Bajari et al. (2008) or Section 5.3), but knowledge of $\hat{\theta}_n^K$ does not allow for point estimation of objects typically of interest to applied researchers, including the evaluation of predicted choice probabilities, marginal effects, price elasticities, and consumer surplus. As we detail in the next section, the SSL, ISL, and SSML models lend themselves to natural lower and upper bounds on such objects of interest, whereas the PMSE cannot say anything about these (because of greater flexibility and robustness in the estimation of the preference parameters θ).

Across both methods and models, which approach to use therefore depends on the nature of the available data and the relative weight placed by researchers on a number of important research design choices. At a high level, we see the choice between the integrating over and differencing out approaches to depend on a researcher's willingness to specify a particular model matching consumers to their choice sets, their ability to integrate over choice sets in estimation, and the importance to their research question of being able to calculate outcomes of interest that depend on information about consumer choice sets. Within the differencing out approach, we see the choice depending again on the nature of the data, but also on the desire for parametric flexibility versus computational ease and the possibility of bounding some objects of interest.

3.5. Unobserved choice sets and market-level data

Most of this paper focuses on approaches using individual-level data, however, we show here with an example that unobserved choice set heterogeneity can also lead to inconsistent estimators in the context of aggregate-level data on market shares (e.g., Goeree, 2008 and Bruno and Vilcassim, 2008).

Consider a parsimonious model of inertia popular in the applied literature (e.g. Ho et al., 2017, Heiss et al., 2016, Hortaçsu et al., 2017, and Abaluck and Adams, 2017). Abaluck and Adams, 2017, in particular, call this model the Default-Specific Consideration (DSC) model. Suppose there are many identical individuals (indexed by i) in a large market, each making choices among $j = 1, \dots, J$ products. We observe the proportion of individuals who choose each of the J products or the outside option $j = 0$ over T time periods: $\{P_{0t}, P_{1t}, \dots, P_{Jt}, \dots, P_{Jt}\}_{t=1}^T$. Denote by $y_{ijt} = 1$ whether individual i purchases product j in period t , $y_{ijt} = 0$ otherwise. The predicted market shares are MNL:

$$\begin{aligned} \Pr[y_{ijt} = 1] &= \frac{\exp(\delta_{jt})}{1 + \sum_{k=1}^J \exp(\delta_{kt})}, \\ &= \frac{\exp(\alpha + X_{jt}\beta + \xi_{jt})}{1 + \sum_{k=1}^J \exp(\alpha + X_{kt}\beta + \xi_{kt})}, \end{aligned} \quad (3.17)$$

where $\delta_t = (\delta_{1t}, \dots, \delta_{Jt})$ are mean utilities (δ_{0t} is normalized to 0), $X_t = (X_{1t}, \dots, X_{Jt})$ and $\xi_t = (\xi_{1t}, \dots, \xi_{Jt})$ are, respectively, product-specific observable and unobservable characteristics, and (α, β) are preference parameters. Whenever every individual is known to make choices from the full choice set $0 \cup \mathcal{J} = \{0, 1, \dots, J\}$, $P_{jt} = \Pr[y_{ijt} = 1]$, $P_{jt} > 0$, and $\mathbb{E}[\xi_{jt} | X_{jt}] = 0$ for each j , then following (Berry, 1994) the model can be consistently estimated by OLS:¹⁸

$$\ln \left(\frac{\Pr[y_{ijt} = 1]}{\Pr[y_{i0t} = 1]} \right) = \alpha + X_{jt}\beta + \xi_{jt}. \quad (3.18)$$

Now suppose that in each t a share $(1 - \rho)$ of individuals do not make any decision among the J "inside" products, but automatically "fall into" the outside option (or any other default option). Conversely, the remaining share ρ of individuals still make choices from the full choice set $0 \cup \mathcal{J}$. This implies that the observed market shares will no longer

¹⁸ To economize on notation, we assume the case in which X_{jt} is exogenous. The intuition carries on to the case in which some of the X_{jt} correlates with ξ_{jt} but the econometrician has valid instruments Z_{jt} .

correspond to the predicted MNL probabilities from (3.17): $P_{jt} = \rho \cdot \Pr[y_{ijt} = 1]$ for $j = 1, \dots, J$ and $t = 1, \dots, T$, and $P_{0t} = (1 + \Delta_t) \cdot \Pr[y_{i0t} = 1]$,¹⁹ where

$$\begin{aligned}\Delta_t &= \frac{(1 - \rho) \cdot \left(\sum_{k=1}^J \Pr[y_{ikt} = 1] \right)}{\Pr[y_{i0t} = 1]} \\ &= (1 - \rho) \cdot \left(\sum_{k=1}^J \exp(\delta_{kt}) \right)\end{aligned}\quad (3.19)$$

captures the increase in the probability of choosing the outside option due to the likelihood of “not making any choice among the inside goods” (this follows from the constraint that the sum of the choice probabilities must be 1). Finally, suppose that $\rho > 0$ so that $P_{jt} > 0$ for each j .

It is easy to see that whenever $\rho < 1$, the standard procedure proposed by Berry (1994) will deliver consistent estimates of (α, β) when T is fixed and $J \rightarrow \infty$. Differently, in the case in which J is fixed and $T \rightarrow \infty$ (see Freyberger, 2015), the OLS estimator will deliver inconsistent estimates. First, when $\rho < 1$, $\ln(P_{jt} \cdot P_{0t}^{-1})$ will differ from (3.18):

$$\ln\left(\frac{P_{jt}}{P_{0t}}\right) = \ln\left(\frac{\rho \cdot \Pr[y_{ijt} = 1]}{(1 + \Delta_t) \cdot \Pr[y_{i0t} = 1]}\right) = \ln\left(\frac{\rho}{1 + \Delta_t}\right) + \alpha + X_{jt}\beta + \xi_{jt}, \quad (3.20)$$

with $\rho \cdot (1 + \Delta_t)^{-1} \in (0, 1)$ and $\ln(\rho \cdot (1 + \Delta_t)^{-1}) < 0$. Suppose $T = 1$ and consider asymptotic approximations in terms of $J \rightarrow \infty$. In this case, i.e. equation (3.20) without the t subscript, by simply regressing $\ln(P_j \cdot P_0^{-1})$ on a 1 and X_j , the resulting OLS estimator will be consistent because $\Delta \rightarrow \infty$ as $J \rightarrow \infty$. This is not surprising, given that as the number of inside products grows large, the relevance of the outside option vanishes, and with it the role played by $\rho < 1$.

Second, consider the case of J fixed and $T \rightarrow \infty$. Regression (3.20) can be re-written as:

$$\begin{aligned}\ln\left(\frac{P_{jt}}{P_{0t}}\right) &= \ln\left(\frac{\rho}{1 + \Delta_t}\right) + \alpha + X_{jt}\beta + \xi_{jt} \\ &= \alpha + X_{jt}\beta + \xi_{jt}^*, \text{ where} \\ \xi_{jt}^* &= \ln\left(\frac{\rho}{1 + \Delta_t}\right) + \xi_{jt}.\end{aligned}\quad (3.21)$$

It is then clear from (3.19) that—whenever $\rho < 1$ —the error term ξ_{jt}^* will be a function of the rest of the model, especially X_t , so that endogeneity will lead to inconsistency of the OLS estimator.²⁰

Most of the existing literature has addressed the problem of unobserved choice set heterogeneity in the context of aggregate-level data on the basis of the integrating over approach (e.g., Goeree, 2008 and Bruno and Vilcassim, 2008). The idea of differencing out does not suit well the case of aggregate-level data; it is not clear how one could use panel data on the evolution of market shares to obtain similar results to those available in the literature for individual-level discrete choice models. In the interest of space, we refer the reader to the excellent surveys by Hickman and Mortimer (2016) and Honka et al. (2019) for an overview of the solutions proposed to address this problem.

4. Further considerations on “differencing out”

Researchers interested in the “integrating over” approach for handling unobserved choice set heterogeneity have a large applied literature on which to rely. For those researchers pursuing the less common “differencing out” approach, we provide further guidance on a number of additional topics.

4.1. Economic foundations of sufficient sets

In this section, we describe a few examples of how choice environments that have been analyzed in a wide variety of literatures in economics map onto particular sufficient sets $f(Y_i)$ that can be used to implement the estimators discussed above.

¹⁹ Equivalently, $P_{0t} = (1 - \rho) + \rho \cdot \Pr[y_{i0t} = 1]$.

²⁰ In this specific example, one can still obtain a consistent OLS estimator of β from a slightly unusual regression equation that excludes observations about the outside option:

$$\ln\left(\frac{P_{jt}}{P_{kt}}\right) = \ln\left(\frac{\rho \cdot \Pr[y_{ijt} = 1]}{\rho \cdot \Pr[y_{ikt} = 1]}\right) = (X_{jt} - X_{kt})\beta + \xi_{jt} - \xi_{kt}. \quad (3.22)$$

4.1.1. Stable choice sets

There are many examples of economic models that give rise to choice settings that are stable over time. [Morgan and Manning \(1985\)](#) present results on the existence and properties of search rules for dynamic search problems in which individuals may choose both the number of periods in which samples of alternatives are searched and the size of the sample searched in each period. The authors show that if individuals have full recall and no lost alternatives, then a *fixed-sample search* strategy is optimal if either the marginal cost of searching or individuals' discount factors are sufficiently high. Such strategies have been studied empirically in both product and labor markets (e.g., [Janssen and Moraga-González, 2004](#), [De los Santos et al., 2012](#)).

Similar factors are at play in [Gaynor et al. \(2016\)](#)'s study of the impact of a regulatory policy that expanded specialists' referral networks and in studies of *school choice*, where the set of schools in any neighborhood usually does not evolve rapidly, and those households that do not change neighborhood face a stable set of schools for their children (e.g., [Walters, 2018](#), [Fack et al., 2019](#)).

Consistent with these three examples, suppose that individuals' choice sets are potentially heterogeneous across i 's but stable over the T choice situations, $CS_{it}^* = CS_i^*$. Let $H_i = \bigcup_{t=1}^T \{Y_{it}\} \subseteq CS_i^*$ be the collection of all the alternatives that individual i is observed to choose in any of the T choice situations. We define the **Full Purchase History (FPH)** sufficient set as $f_{FPH}(Y_i) = (H_i)^T$, the set of choice sequences given by the cartesian product of H_i in each of the T choice situations. Note that $f_{FPH}(Y_i)$ implies a **FPH SSL** model like that of the SSL, ISSL, and SSML models described in (3.6) and (3.7) which are simple to implement.

In addition to the FPH sufficient set, the assumption of stable choice sets also underpins another sufficient set: that proposed by [Chamberlain \(1980\)](#) for the classic Fixed Effect logit model (FE logit). In a model with systematic utilities given by $V_i(X_{ijt}, \theta) = \delta_{ijt} + X_{ijt}\beta$, [Chamberlain \(1980\)](#) shows that β can be consistently estimated by the ML estimator of a SSL model with sufficient set $f_{CP}(Y_i) = \mathcal{P}(Y_i)$: the set of all possible permutations of observed choice sequence Y_i . As such, we call this the **Choice Permutations (CP)** sufficient set and the corresponding model the **CP SSL**.²¹

[Chamberlain \(1980\)](#)'s expressed motivation for the sufficient set $f_{CP}(Y_i) = \mathcal{P}(Y_i)$ was to difference out the fixed effects (δ_{ijt}) from each individual's systematic utility. But his assumption of choice set stability also implies that $f_{CP}(Y_i) \subseteq CS_i^*$. As such, sufficient set $f_{CP}(Y_i) = \mathcal{P}(Y_i)$ will not only accommodate unobserved preference heterogeneity in the form of individual-alternative specific fixed effects, but also unobserved choice set heterogeneity.²² [Matejka and McKay \(2015\)](#) propose a choice model with rational inattention in which the reduced form choice probabilities take the form of a CP SSL model (see Theorem 1, p.282).

As is well known, the CP SSL does not usually allow the identification of i 's fixed effects δ_{ijt} 's, but only those elements of β associated with time-varying observables. This can limit its usefulness to applied researchers. By contrast, the SSL and SSML models obtained from $f_{FPH}(Y_i)$, while relying on the same assumption of choice set stability, typically allow the identification of all parameters.

4.1.2. Growing choice sets

The possibility that choice sets may grow over time is a consequence of many search models. For example, in addition to the results in the previous section, [Morgan and Manning \(1985\)](#) also show that, if the assumptions ensuring full recall and 'no lost alternatives' hold, then any *sequential search* strategy over T periods will imply choice sets that are weakly growing over time, so that $CS_{it}^* \subseteq CS_{it+1}^*$. Similarly, [Caplin and Dean \(2011\)](#) propose two models of sequential search: an alternative-based search model, which provides the micro foundations underlying some of the functional form restrictions used by [Goeree \(2008\)](#), [Manzini and Mariotti \(2014\)](#), and [Abaluck and Adams \(2017\)](#) to aid the identification of [Manski \(1977\)](#)'s model, and a reservation-based search model which is a formalization of [Simon \(1955\)](#)'s satisficing model. In related work, [Caplin et al. \(2011\)](#) find experimental evidence in support of this latter model.

[Masatlioglu and Nakajima \(2013\)](#) propose another dynamic search framework that they call Choice by Iterative Search, which also implies weakly growing choice sets. Several models in the fast-growing literature on limited attention build on this framework. An example is [Eliaz and Spiegel \(2011\)](#), who study a setting in which individuals have a singleton status quo, i.e. a choice set including only one product (possibly different across individuals), and firms seek to use marketing devices, e.g. advertising, to include their products in individuals' choice sets.²³ A dynamic extension of this framework, in which multiple firms compete in each period with advertising to encourage individuals to consider their products and

²¹ Note that the CP sufficient set, $f_{CP}(Y_i)$, cannot be expressed as the cartesian product of t -specific sufficient sets, giving rise to models that are harder to implement (e.g., the CP SSL model can be expressed as in (3.5) but not as in (3.6)). For those cases where T is large and/or there is substantial heterogeneity in the alternatives chosen across the T choice situations, the computational burden implied by the CP SSL can be considerable. [D'Haultfœuille and Iaria \(2016\)](#) show how to ease this computational burden by applying the insights of [McFadden \(1978\)](#) to the estimation of β from (uniform) random subsets of $f_{CP}(Y_i)$.

²² The argument is essentially identical to that leading to (3.5), except for the different systematic utilities that in [Chamberlain \(1980\)](#) have individual-alternative specific coefficients. By replacing $V(X_{ijt}, \theta)$ with $\delta_{ijt} + X_{ijt}\beta$ and $f(Y_i)$ with $\mathcal{P}(Y_i)$ in Eq. (3.5), [Chamberlain \(1980\)](#)'s result follows.

²³ Other relevant examples of the application of this framework are [Ho et al. \(2017\)](#) and [Heiss et al. \(2016\)](#), which study the Medicare Part D program and document that individuals switch health plans infrequently and search imperfectly, possibly because of high search costs.

status quos evolve over time would also yield weakly increasing choice sets; this motivates the approach in our empirical application in Section 5.

Other models beyond search can imply growing choice sets as well. Kőszegi and Szeidl (2013), analyze the impact of “focus” on individuals’ choices, providing numerous examples of individuals focusing on one of an alternative’s (possibly many) attributes, leading them to select an alternative that exceeds others in this attribute, even if a comparison of the alternatives across all attributes would lead to a different choice. As they describe themselves, “Formally, there are T periods and in period t , a consumer makes a choice x_t from the deterministic ... consideration set $X_t(h_{t-1})$, where $h_{t-1} = (x_1, \dots, x_{t-1})$ is the history of choices up to period $t - 1$ ”. In a repeat-purchase environment (e.g., retail purchases of household goods), h_{t-1} would consist of the history of that individual’s previous purchase decisions, a fact that can be used to form a sufficient set as we describe next.

These three examples suggest the following sufficient set. Let $H_{it} = \bigcup_{b=1}^t \{Y_{ib}\} \subseteq CS_{it}^*$ be the collection of all the alternatives that individual i is observed to choose between choice situation 1 and t . We define the **Past Purchase History (PPH)** sufficient set as $f_{PPH}(Y_i) = \times_{t=1}^T H_{it}$, the cartesian product of H_{it} between choice situation 1 and T . Note that, similarly to the Full Purchase History sufficient set, $f_{PPH}(Y_i)$ implies a **PPH SSL** model like that of the SSL, ISSL, and SSML models described in (3.6) and (3.7), which are very simple to implement. As with the FPH sufficient set, the intuition is to exploit the variation in the characteristics of only these alternatives over time.

4.1.3. Inter-personal comparisons

While the primary focus of sufficient sets in this paper is their use in panel data environments, they can also be constructed in cross-sectional environments to a group of individuals, each making a separate purchase decision at a single point in time as long as they purchase from the same choice set.²⁴ This could be possible, for example, for the question of whether greater availability of fast food outlets causes obesity as in Currie et al. (2010). The authors collect precise geographic data on the location of fast food outlets and where children live and attend school and examine the effect of the presence of a fast food restaurant within given distances of the school attended by the student. If the authors were willing to assume that all children living on the same street and attending the same school faced the same choice set, they could conclude that all such outlets were in the choice set for all such children and this could form the basis for a sufficient set in our approach.

In such settings, one can call each i a “consumer type” and each t one of the T individuals of that type.²⁵ Then, when the same choice set is faced by the T individuals of the same consumer type i , the econometrician can use the **Inter-Personal** sufficient set: $f_{IP}(Y_i) = (H_i)^T$, where $H_i = \bigcup_{t=1}^T \{Y_{it}\} \subseteq CS_{it}^*$.²⁶ The sufficient set $f_{IP}(Y_i)$ imputes to each individual t the collection of all the alternatives observed to be chosen by any of the T individuals of consumer type i . As for $f_{PPH}(Y_i)$ and $f_{FPH}(Y_i)$, note that $f_{IP}(Y_i)$ also implies a **IP SSL** model like that of the SSL, ISSL, and SSML models described by (3.6) and (3.7), which are very simple to implement.

4.2. Bounding functions of the preference parameters

The various approaches presented in Section 3 differ in the extent to which they allow us to evaluate functions of θ (or θ_i for the ISSL model), for example: willingness to pay, elasticities, consumer surplus, or the analysis of counterfactuals, such as evaluating the effects of a change in tax policy or a merger between manufacturers. At one extreme is Fox (2007)’s PMSE discussed in sub Section 3.2.3. The PMSE can point-estimate the preference parameters θ , as well as simple functions of these (e.g., willingness-to-pay), but cannot reveal functions of θ that involve knowledge of the distribution of the unobserved portion of utility.²⁷ At the other extreme are models based on Manski (1977) that integrate over unobserved choice sets discussed in subSections 3.1 and 3.3. These approaches involve specifying a model of choice set formation that reveals the distribution of choice sets in the population, allowing the point-identification of all functions of θ that depend on this distribution.

The SSL, the ISSL, and the SSML represent an intermediate case. Here we describe parameters and functions of parameters that we can point-identify, and how we can use sufficient sets to derive bounds on several useful functions of these parameters. For ease of notation, we limit the discussion to the SSL with the understanding that similar ideas readily apply also to the ISSL and to the SSML. To simplify exposition, suppose that the systematic utilities take the

²⁴ Note the validity of this sufficient set further relies on the assumption that the observable characteristics of any product j are the same for each of the T individuals of consumer type i , or that the econometrician knows how they change across t ’s.

²⁵ This is without loss of generality. We could allow each type to have a different number of individuals, T_i , but this would only complicate notation and provide no deeper insights into the underlying mechanism at work.

²⁶ Note that this definition of $f_{IP}(Y_i)$ is identical to that for the Full Purchase History sufficient set, but because the underlying economic environments are so different (e.g. i is an individual and t is a time period in $f_{PPH}(Y_i)$, while i is a consumer type and t is an individual in $f_{FPH}(Y_i)$), we prefer to define the two separately.

²⁷ See Fox (2007) and Bajari et al. (2008) for more details on this point.

form:

$$V(X_{ijt}, \theta) = \delta_{jt} + X_{ijt}\beta + \alpha p_{jt},$$

where $\theta = [\delta_1, \dots, \delta_J, \beta, \alpha]$ and p_{jt} is the price of alternative j_t . We can point-identify the vector of preference parameters θ from the SSL model $\Pr[Y_i = j|f(Y_i) = r_i, \theta]$.²⁸ We can similarly point-identify simple functions of θ . For example, we are often interested in willingness-to-pay (WTP) for product characteristic k , X_{ijt}^k . By Roy's Identity, this can be computed as:

$$WTP_k = -\frac{\partial V_{ijt}/\partial X_{ijt}^k}{\partial V_{ijt}/\partial p_{jt}} = -\frac{\beta_k}{\alpha}. \quad (4.1)$$

Other outputs of economic interest, however, require information about the distribution of choice sets in the population for point-identification. We cannot point-identify these functions, but we can place bounds on them. This makes clear that point-identification relies on strong assumptions about the choice set formation process. The probability with which i chooses alternative j_t given choice set $CS_{it}^* = c_{it}$ is

$$Pr_{ijt}^{CS^*}(\theta) \equiv \Pr[Y_{it} = j_t | CS_{it}^* = c_{it}, \theta] = \frac{\exp(\delta_{jt} + X_{ijt}\beta + \alpha p_{jt})}{\sum_{m \in c_{it}} \exp(\delta_m + X_{imt}\beta + \alpha p_{mt})}, \quad (4.2)$$

if $j_t \in CS_{it}^* = c_{it}$ and zero otherwise. This choice probability depends on i 's (unobserved) choice set, CS_{it}^* . Suppose that we observe a superset Q_{it} of the true but unobserved choice set, so that $CS_{it}^* \subseteq Q_{it}$. This could be, for example, the collection of all alternatives observed to be chosen by any i in choice situation t . It follows that, even if we do not directly observe $CS_{it}^* = c_{it}$, $f_t(Y_i) \subseteq CS_{it}^*$ and $CS_{it}^* \subseteq Q_{it}$. We can therefore use these conditions to bound the true but unobserved denominator of the SSL choice probabilities for any $X_{it} = [X_{i1t}, p_{1t}, \dots, X_{ijt}, p_{jt}]$ and θ :

$$\sum_{m \in f_t(Y_i)=r_{it}} \exp(\delta_m + X_{imt}\beta + \alpha p_{mt}) \leq \sum_{m \in CS_{it}^*=c_{it}} \exp(\delta_m + X_{imt}\beta + \alpha p_{mt}) \leq \sum_{m \in Q_{it}=q_{it}} \exp(\delta_m + X_{imt}\beta + \alpha p_{mt}). \quad (4.3)$$

Denote for brevity also $Pr_{ijt}^Q(\theta) \equiv \Pr[Y_{it} = j_t | Q_{it} = q_{it}, \theta]$ and $Pr_{ijt}^f(\theta) \equiv \Pr[Y_{it} = j_t | f_t(Y_i) = r_{it}, \theta]$. It follows from (4.3) that for any $j_t \in f_t(Y_i) = r_{it}$:

$$Pr_{ijt}^Q(\theta) \leq Pr_{ijt}^{CS^*}(\theta) \leq Pr_{ijt}^f(\theta). \quad (4.4)$$

That is to say, the true choice probability with which i chooses j_t in t is bounded from below by the same probability assuming i chooses from some superset of the unobserved choice set, $Q_{it} = q_{it}$, and from above by the same probability assuming i chooses from just their sufficient set, $f_t(Y_i) = r_{it}$. Observe that $Pr_{ijt}^f(\theta)$ takes the usual logit form whenever $j_t \in r_{it}$, but that it equals zero whenever $j_t \notin r_{it}$. Hence, for those $j_t \in q_{it}$ but $j_t \notin r_{it}$, $Pr_{ijt}^f(\theta)$ will not be a valid upper bound for $Pr_{ijt}^{CS^*}(\theta)$: even if $j_t \notin r_{it}$, it can still be the case that $j_t \in CS_{it}^* = c_{it}$ and so that $Pr_{ijt}^{CS^*}(\theta) > 0$. Similarly, among the $j_t \in q_{it}$ that $j_t \notin r_{it}$, there can be some $j_t \in c_{it}$. But for those $j_t \in q_{it}$ that $j_t \notin c_{it}$, $Pr_{ijt}^Q(\theta) > Pr_{ijt}^{CS^*}(\theta) = 0$: $Pr_{ijt}^Q(\theta)$ will not be a valid lower bound for $Pr_{ijt}^{CS^*}(\theta)$. It is then unclear how to bound $Pr_{ijt}^{CS^*}(\theta)$ for those $j_t \in q_{it}$ but $j_t \notin r_{it}$. However, it is always possible to construct bounds for the probability with which i would choose j_t if indeed j_t were to be added to their true but unobserved choice set, $CS_{it}^* \cup \{j_t\} = c_{it} \cup \{j_t\}$:

$$Pr_{ijt}^{CS^* \cup j_t}(\theta) \equiv \Pr[Y_{it} = j_t | CS_{it}^* \cup \{j_t\} = c_{it} \cup \{j_t\}, \theta] = \frac{\exp(\delta_{jt} + X_{ijt}\beta)}{\sum_{m \in c_{it} \cup \{j_t\}} \exp(\delta_m + X_{imt}\beta)}. \quad (4.5)$$

By defining $Pr_{ijt}^{Q \cup j_t}(\theta)$ and $Pr_{ijt}^{f \cup j_t}(\theta)$ analogously, note that $Pr_{ijt}^{Q \cup j_t}(\theta) = Pr_{ijt}^Q(\theta)$, $Pr_{ijt}^{CS^* \cup j_t}(\theta) = Pr_{ijt}^{CS^*}(\theta)$, and $Pr_{ijt}^{f \cup j_t}(\theta) = Pr_{ijt}^f(\theta)$ for any $j_t \in r_{it}$, while $Pr_{ijt}^{Q \cup j_t}(\theta) \leq Pr_{ijt}^{CS^* \cup j_t}(\theta)$ and $Pr_{ijt}^{CS^* \cup j_t}(\theta) \leq Pr_{ijt}^{f \cup j_t}(\theta)$ for any $j_t \notin r_{it}$. Using these facts, we can then complement condition (4.4) for those $j_t \notin r_{it}$ and propose choice probability bounds for all (i, j_t, t) combinations:

$$Pr_{ijt}^{Q \cup j_t}(\theta) \leq Pr_{ijt}^{CS^* \cup j_t}(\theta) \leq Pr_{ijt}^{f \cup j_t}(\theta). \quad (4.6)$$

Condition (4.6) can be used to construct bounds for functions of individual choice probabilities, such as average choice probabilities or elasticities. The average choice probability of alternative j_t for a certain group of individuals $i = 1, \dots, I_t$ can be bounded by:

$$I_t^{-1} \sum_{i=1}^{I_t} Pr_{ijt}^{Q \cup j_t}(\theta) \leq I_t^{-1} \sum_{i=1}^{I_t} Pr_{ijt}^{CS^* \cup j_t}(\theta) \leq I_t^{-1} \sum_{i=1}^{I_t} Pr_{ijt}^{f \cup j_t}(\theta). \quad (4.7)$$

²⁸ Note that here, differently from most other parts in the paper, we will keep track of the " i " subscript in the realizations of the sufficient sets, $f(Y_i) = r_i$, and of the choice sets, $CS_{it}^* = c_{it}$. This is essential to avoid confusion when computing averages across individuals, as detailed below.

With indirect utilities that are linear in price, individual i 's own- and cross-price elasticities are:

$$\begin{aligned}\xi_{it}^{ij}(X_{it}, \theta) &= \beta_p p_{j,t} (1 - Pr_{ijt}^{CS^* \cup j}(\theta)) \\ &= \beta_p p_{j,t} \left(1 - \frac{\exp(\delta_{jt} + X_{ijt} \beta)}{\sum_{m \in c_{it} \cup \{j_t\}} \exp(\delta_m + X_{imt} \beta)} \right), \\ \xi_{it}^{jk}(X_{it}, \theta) &= \beta_p p_{k,t} Pr_{ikt}^{CS^* \cup j}(\theta) \\ &= -\beta_p p_{k,t} \left(\frac{\exp(\delta_{kt} + X_{ikt} \beta)}{\sum_{m \in c_{it} \cup \{j_t\}} \exp(\delta_m + X_{imt} \beta)} \right),\end{aligned}\tag{4.8}$$

where $p_{j,t}$ is j_t 's price in choice situation t and β_p is the price coefficient. As (4.8) makes clear, even though we may have a consistent estimator of $\delta = [\delta_1, \dots, \delta_j, \dots, \delta_J]$ and β , we still do not know the exact $CS_{it}^* \cup \{j_t\} = c_{it} \cup \{j_t\}$ for each i and t , and thus the true $Pr_{ijt}^{CS^* \cup j}(\theta)$, $\forall j_t \in CS_{it}^* \cup \{j_t\} = c_{it} \cup \{j_t\}$. Given (4.8), (4.6), and $\beta_p < 0$, we obtain the following bounds on the elasticities for any j_t , k_t , X_{it} , δ , and β :

$$\begin{aligned}\underbrace{\beta_p p_{j,t} (1 - Pr_{ijt}^{f \cup j}(\theta))}_{\text{Lower (in abs. value) Bound}} &\leq \xi_{it}^{ij}(X_{it}, \theta) \leq \underbrace{\beta_p p_{j,t} (1 - Pr_{ijt}^{Q \cup j}(\theta))}_{\text{Upper (in abs. value) Bound}} \\ \underbrace{-\beta_p p_{k,t} Pr_{ikt}^{Q \cup j}(\theta)}_{\text{Lower Bound}} &\leq \xi_{it}^{jk}(X_{it}, \theta) \leq \underbrace{-\beta_p p_{k,t} Pr_{ikt}^{f \cup j}(\theta)}_{\text{Upper Bound}}.\end{aligned}\tag{4.9}$$

The same bounds in Eq. (4.3) imply the ability to bound consumer surplus. Let the true consumer surplus of individual i in t be:

$$W_{it}(X_{it} | \theta, CS_{it}^* = c_{it}) = \zeta + \frac{1}{\alpha} \ln \left(\sum_{m \in c_{it}} \exp(\delta_m + X_{imt} \beta + \alpha p_{mt}) \right),\tag{4.10}$$

where ζ is Euler's constant. Then, for any X_{it} and θ :

$$W_{it}(X_{it} | \theta, f_t(Y_i) = r_{it}) \leq W_{it}(X_{it} | \theta, CS_{it}^* = c_{it}) \leq W_{it}(X_{it} | \theta, Q_t = q_{it}).\tag{4.11}$$

As an example of how to conduct inference on the identification regions described in this section, in Appendix E we provide confidence intervals for the elasticity bounds on the basis of Imbens and Manski (2004).

Partial identification. Some recent working papers take a different stand from the approaches surveyed here and address the problem of unobserved choice set heterogeneity on the basis of partial identification methods. For example, Lu (2018) proposes conditions for both partial and point-identification of discrete choice models. The main conditions required for partial identification are knowledge of the smallest and the largest possible choice sets, and a monotonicity property derived from utility maximization (i.e., Sens's α property). Barseghyan et al. (2019) further weaken the requirements for partial identification and characterize the sharp identification region from knowledge of the minimum size of the true but unobserved choice sets. Both papers stress the trade-offs between identification power and several of the assumptions commonly used in the methods described in this survey. They illustrate that interesting features of the model can be identified even when unobserved choice sets are endogenously matched to individuals on the basis of their preferences.

4.3. Specification tests: Choice set stability and IIA

The correct implementation of the differencing out approach relies on two kinds of assumptions: assumptions about the evolution of choice sets across choice situations and assumptions about unobserved preference heterogeneity. In Section 4.1, we discussed examples of choice set formation processes giving rise to sufficient sets compatible with Conditions 1 and 2, and in Section 3.2 different models that rely on the IIA property to different extents. In what follows, we illustrate how existing testing procedures can be used to discriminate among alternative choice set formation processes and various departures from the IIA property.

4.3.1. Testing among competing sufficient sets

In the context of the ML estimator of the SSL and the ISSL models, alternative sufficient sets lead to more or less robust and/or efficient estimators along the lines of Hausman and McFadden (1984) and can be used to form specification tests. We discuss here how to test for some of the assumptions implicit in several sufficient sets, such as the length of the sequence of choice situations for which choice sets are stable or grow.

The basis for these specification tests is the *Factorization Theorem* proposed by Ruud (1984) and further explored by Hausman and Ruud (1987). Ruud (1984)'s result enables one to rank Maximum Likelihood Estimators (MLEs) of SSL or ISSL models with different sufficient sets in terms of their efficiency: the MLE of a SSL or ISSL model with sufficient set f_L is more efficient than the MLE of a SSL or ISSL model with sufficient set $f_Z \subset f_L$. This result can be applied recursively, so that if two subsets of f_L are available, say f_Z and f_{XZ} with $f_{XZ} \subset f_Z$, then the efficiency rank of the three MLEs will be $f_L \succ f_Z \succ f_{XZ}$. As we detail in Appendix F, building on the Factorization Theorem one can construct Hausman tests between SSL or ISSL models based on different sufficient sets and implicitly test for underlying economic assumptions such as choice set stability or the IIA property. For example, in the context of the SSL model, the sufficient sets discussed earlier rely on the following economic assumptions:

- f_{CP} : Choice set stability across T choice situations and the possibility of IIA violations in the form of individual-alternative specific fixed effects, δ_{it} .
- f_{FPH} : Choice set stability across T choice situations and the IIA property.
- f_{PPH} : Choice set evolution in the form of weakly growing choice sets (or, symmetrically, weakly shrinking choice sets) across T choice situations and the IIA property.²⁹

The first possibility is to compare f_{CP} , f_{FPH} , and f_{PPH} for choice sequences of constant length T . In this case, both the CP and PPH sufficient sets are subsets of the FPH sufficient set: $f_{CP}(Y_i) \subset f_{FPH}(Y_i)$ and $f_{PPH}(Y_i) \subset f_{FPH}(Y_i)$ for any $Y_i \in \mathcal{CS}_i^* = c$. As we discuss in Appendix F.1, these relationships can be used to test for the assumption of choice set stability and for violations of the IIA property.

The second possibility is to fix a specific f , say f_{CP} , and to compare choice sequences with some of their sub-sequences: for example, the sequence $1, 2, \dots, T^L$ can be split into two mutually exclusive sub-sequences $1, 2, \dots, T^Z$ and $T^Z + 1, \dots, T^L$, and this gives rise to different f_{CP} 's, f_{CP}^Z (separately from 1 to T^Z and from $T^Z + 1$ to T^L) and f_{CP}^L (from 1 to T^L) such that $f_{CP}^Z(Y_i) \subset f_{CP}^L(Y_i)$ for any $Y_i \in \mathcal{CS}_i^* = c$. The same holds both for f_{FPH} and f_{PPH} . As illustrated in Appendix F.1, these comparisons allow one to test for general forms of choice set stability or evolution.

4.3.2. Testing for departures from the IIA

A classic simple test for the IIA property proposed by McFadden et al. (1977) involves a comparison between a MNL with its *true* choice set against another MNL with a *restricted* choice set. The specification test described in the previous subsection is based on the same logic but in a more complex environment where true choice sets are not observed. The additional layer of complexity leads to some ambiguity in the classic testing procedure, because rejection of the null can now be motivated by either a failure of the IIA property (as in Hausman and McFadden, 1984) or by the sufficient sets being too large (a violation of Condition 1), or by both simultaneously. As illustrated in Section 2.2, the imputation in estimation of a choice set that is too "large" (so that Condition 1 does not hold) is mechanically equivalent to a violation of the IIA property. In general, such ambiguity cannot be fully resolved: any testing procedure of this kind will be valid and informative only under some maintained assumptions. At a deeper level, this is a fundamental identification problem: as discussed by McFadden (1987), any discrete choice model can be formally re-written as a model satisfying the IIA property, but with a complex dependence on the explanatory variables. We illustrate some examples of maintained assumptions necessary for the test to be valid in Appendix F.1. For instance, under the maintained assumption of choice set stability and a specific alternative model of unobserved preferences (i.e., Gumbel errors plus individual-alternative specific fixed effects), one can test for departures from the IIA property by comparing the estimates of a CP SSL versus those of a FPH SSL. Both the CP sufficient set and the FPH sufficient set require choice sets to be stable, but – differently from the FPH SSL – the CP SSL controls for individual-alternative specific fixed effects which may induce violations of the IIA.

A second way to test for departures of the IIA can be based on the more general models discussed in Section 3.2.2. For any given correctly specified sufficient set, in the ISSL model (3.7) one can check whether the I estimates $\hat{\theta}_i$'s are statistically indistinguishable across individuals. Similarly, in the SSML model (3.12), for any given correctly specified $f(Y_i) = r$, one can check whether Ψ^r is degenerate, i.e. all but one of the Q probability mass functions ψ_q^r 's are equal to zero. A third tool that can be used to further investigate failures of the IIA property is the nested logit version of the testing procedure proposed by Hausman and McFadden (1984), which consists of comparing a nested logit against a MNL, both from the *true* choice set. In Appendix G, we illustrate how under similar assumptions to those required by the MNL, sufficient sets can also be used for the consistent estimation of nested logit models with unobserved choice set heterogeneity. In particular, for any given correctly specified $f(Y_i) = r$, it is possible to consistently estimate the *within-nest* part of a nested logit model at very little additional cost with respect to a MNL model, and this is enough to implement a test for departures of the IIA along the lines of Hausman and McFadden (1984) in the context of unobserved choice set heterogeneity.

²⁹ Importantly, the IIA requirement follows from the SSL model and it is not intrinsic in the f_{FPH} and f_{PPH} sufficient sets. Neither sufficient set relies on the IIA property across individuals when employed in the ISSL model, or even within individuals in the SSML model (see Section 3.2.2).

5. An empirical illustration

There are many empirical examples of the “integrating over” approach, but relatively few of the “differencing out” method. To show how the ideas presented in this survey can be applied in practice, we therefore present an empirical illustration of the differencing out approach. In Section 4.1.2, we discussed models of limited attention and the role that marketing expenditure can play at influencing consumers’ choice sets (as in the models of Eliaz and Spiegel, 2011 and Goeree, 2008). We use data and methods similar to those in Dubois et al. (2020) to estimate demand for chocolate bars by a sample of adult women in the UK making decisions on-the-go, i.e. chocolate purchased outside of the home in small corner stores, vending machines, concession stands, and other outlets for immediate consumption. In the language of Section 3.2.2, we will estimate an ISSL model.

We are interested in estimating consumers’ responsiveness to price and how advertising might affect consumers’ choices. Advertising is important in the chocolate market, and there is intuitive appeal to the idea that ads might play an important role both in bringing products to consumers’ attention (as in Eliaz and Spiegel, 2011 and Goeree, 2008) as well as potentially entering their utility directly (as in Becker and Murphy, 1993).

At any point in time there are more than 100 products available to choose from. In such a choice environment, it is unlikely that an individual will spend the time to consider each one, and collecting information on which products the individual considered (for example, using eye-tracking technologies) is expensive. We compare results from estimation based on the **Complete** sufficient set – where we assume that each individual considers all of the products that are available in the type of store in which they are currently shopping – with the **Past Purchase History (PPH)** sufficient set. We allow for the possibility that individuals have finite memory of products that they have purchased, and consider sufficient sets based on purchase histories of shorter duration (described below). For brevity we omit the “store-type specific” modifier from each of these descriptions.

5.1. Model

We adapt the general ISSL model presented in Section 3.2.2 to the demand for chocolate on-the-go and our data. We assume that each individual makes a purchase from their own (unobserved) choice set CS_{it}^* . It could include many or only a few of the products currently available in the market; it always includes the option not to purchase. We observe what product was purchased, the price paid, the type of store the product was purchased in, and what products have been purchased by others (so are available) in that type of store.

We rely on the large number of choice situations observed per individual to specify choice probabilities as in the ISSL model (3.7). The probability with which individual i buys the sequence of products $j = (j_1, \dots, j_t, \dots, j_T)$ given her sufficient set $f(Y_i) = r_i$ is:

$$\Pr[Y_i = j | f(Y_i) = r_i, \theta_i] = \prod_{t=1}^T \frac{\exp(V(X_{ijt}, \theta_i))}{\sum_{k \in r_{it}} \exp(V(X_{ikt}, \theta_i))}, \quad (5.1)$$

where $f(Y_i) = \times_{t=1}^T f_t(Y_i) = \times_{t=1}^T r_{it}$ and each r_{it} is the set of chocolate bars belonging to individual i ’s sufficient set in week t . Utility for any chocolate bar j_t in week t is given by

$$U_{ijt} = V(X_{ijt}, \theta_i) + \epsilon_{ijt}, \quad (5.2)$$

with

$$V(X_{ijt}, \theta_i) = \delta_{gb} + \alpha_i p_{ojt} + \beta_g \ln(a_{ibt}), \quad (5.3)$$

where δ_{gb} is a brand (to which product j_t belongs) fixed effect for demographic group g , p_{ojt} is the price of product j_t in store-type o in week t , and $\ln(a_{ibt})$ is log advertising exposure to brand b in week t .³⁰ The price variable and our measure of advertising exposure are defined in the next subsection. Each individual has her own price sensitivity, α_i , with a common brand and advertising sensitivity according to their membership in one of nine demographic groups defined by age and equivalised income, which are indexed by g .

The utility of the outside option of not purchasing a chocolate bar is given by

$$U_{i0t} = \delta_{g0} + \sum_m \tau_{gm} + \epsilon_{i0t}, \quad (5.4)$$

where the τ_{gm} ’s are demographic-group-specific month effects meant to capture seasonality and/or cyclicity in on-the-go chocolate demand.

³⁰ We specify brand dummies for eight large chocolate brands.

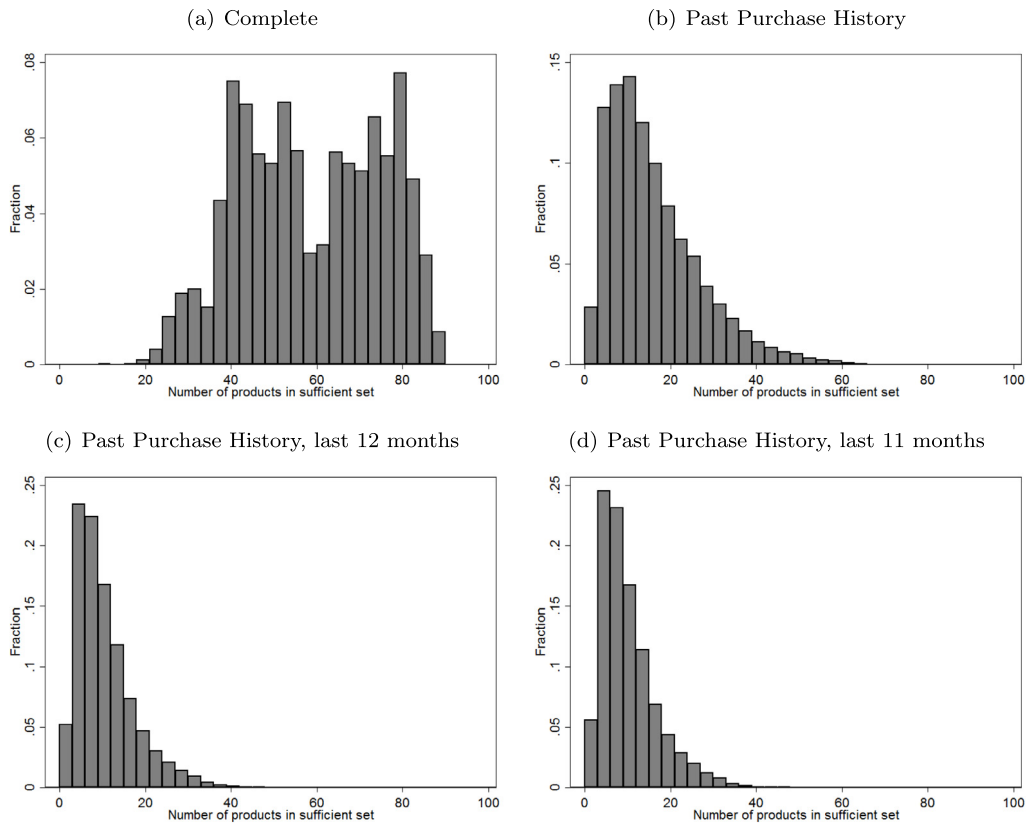


Fig. 1. Number of products in sufficient sets. Note: The histograms shows the distribution of the number of products in the sufficient sets across all purchase occasions.

5.2. Data

We use data on 532 women. The data are from the Kantar Worldpanel on-the-go survey, collected from individuals who record purchases that they make on-the-go for immediate consumption.³¹ We use information on 130,304 purchase occasions over the period 2009–2014. A purchase occasion is when a woman is observed purchasing a snack of any form on-the-go.

At any one point in time, there are more than 100 different types of chocolate products available in the market. The outside option, when a chocolate bar is not purchased, has a 39.6% market share. The three largest market share products are KitKat, with a market share of 3.7%, Cadbury's Twirl, 2.7%, and Cadbury's Dairy Milk, 2.5%.

Individuals purchase products in different outlets. We consider four types of outlets: large national chains (30.1% of sales), news agents (25.2% of sales), vending machines (5.3% of sales), and other types of small stores and outlets (38.6% of sales). We assume that the outlet that we observe the individual shopping in is chosen independently from demand shocks for any specific product. Prices are constructed at the level of the store-type o and week t . We observe prices on each individual transaction and aggregate them to the level of the outlet and week (using the median); most national chains in the UK price nationally, we allow prices in news agents and other outlets to vary across broad regions. 95% of prices range from 20 pence to £1.00, with a few exceptional items available at very low price (for example, Cadbury's Dairy Milk Buttons for 19 pence) and a few large items (for example, a 360g Toblerone Milk Chocolate bar for £4.99).

Fig. 1 shows the distribution of sizes of the sufficient sets used in estimation; panel (a) shows the distribution of the number of chocolate bars in the Complete sufficient set across all purchase occasions. The distribution is bi-modal, with sufficient sets when purchasing from national outlets populating the right mode (up to a maximum of 90 chocolate bars)

³¹ These data were used to analyze the effects of banning advertising in the market for junk foods in Dubois et al. (2018) and in Dubois et al. (2020) to study the impact of soda taxes; we follow their lead in many aspects of our data construction.

Table 3
Coefficient estimates.

		Complete	PPH	PPH 12 months	PPH 11 months
Price	Mean	−1.777	−0.921	−0.643	−0.623
	Std Dev	5.154	1.822	1.759	1.845
Advertising	Mean	0.168	0.054	0.033	0.031
	Std Dev	0.047	0.036	0.024	0.022
Product effects		Yes	Yes	Yes	Yes
Time effects		Yes	Yes	Yes	Yes

Notes: The Table describes the mean and standard deviation of the 532 individual price and 532 advertising parameters. Each column shows the estimates using the indicated sufficient set.

and sufficient sets when purchasing from a vending machine populating the bulk of the left tail. Panel (b) shows the distribution of the number of chocolate bars in Past Purchase History (PPH) sufficient sets; these range from 2 to 64. Panel (c) shows the distribution for the Past Purchase History using only purchases made in the 12 months prior to the current choice occasion, and panel (d) using only those made in the 11 months prior to the current choice occasion—this reduces the sufficient sets to a maximum of 48 products.

To measure advertising exposure we convert weekly advertising (“flows”) into an advertising “stock;” advertising stocks are the depreciated accumulation of the flows. We use minutes of TV advertising to define advertising flows. Following Goeree (2008), we measure advertising exposure at the *individual* level. We use detailed information about when individual ads were aired on television matched with self-reported viewing information. We denote the stock of advertising $stock_{ibt}$: $stock_{ibt}$ ranges from 0 for individuals that do not watch TV, or only watch advertising-free public TV (the BBC), to over 100 min of accumulated exposure to advertisements for a particular brand. The mean is 10 min of accumulated exposure. We follow Dubé et al. (2005) and allow for diminishing returns to advertising by transforming the stock of advertising, $stock_{ibt}$, using the log inverse hyperbolic sine function, $\ln(a_{ibt}) = \ln\left(stock_{ibt} + \sqrt{1 + stock_{ibt}^2}\right)$. Further details on the data and our definitions are available in Appendix H.

5.3. Coefficient estimates

Table 3 presents the mean and standard deviation of the estimated price and advertising coefficients using each of the four sufficient sets.³² The mean of the coefficient on price reduces substantially from the Complete sufficient set to the Past Purchase History, and reduces again when we use only information on purchases made in the year prior to the current choice occasion; restricting to using only the past 11 months does not substantially change the mean. The standard deviation of the individual estimates is smaller for the estimates using the Past Purchase History. Similarly, for the advertising coefficients, the mean of the estimates is higher when using the Complete sufficient set than when using the Past Purchase Histories.

Fig. 2 shows the distribution of the 532 estimated price coefficients across the four sufficient sets. For any individual i , the Complete sufficient set is a superset of the full PPH and the full PPH is a superset of the 12 month’ PPH, which in turn is a superset of the 11 months PPH sufficient set. It is evident that assumptions on individuals’ choice sets have an impact on these estimated distributions.

We perform some of the Hausman tests discussed in Section 4.3.1, reported in Table 4. We report the distribution of p-values of a Hausman test on the price coefficient and separately on the advertising coefficient for each individual; the validity of these relies on the maintained assumptions that unobserved preference heterogeneity is correctly specified by ISSL model (5.1) and that the smallest of the sufficient sets used as a reference is small enough to satisfy Condition 1. Overall, these results suggest that both the Complete and the PPH may be too large and systematically include products not considered or unavailable to individuals when making choices on-the-go. Among the proposed sufficient sets, the most robust – in the sense of Condition 1 – is the PPH using 11 months of previous purchases. Consequently, we regard the comparison between the PPH 12 months versus the PPH 11 months as the most informative: for a substantial share of the sample (37% in the top panel and 63% in the bottom panel), the Hausman tests provide some evidence that, during any purchase occasion, individuals consider at least the chocolate bars they bought in the previous year. In general, if a researcher is not satisfied by the frequency with which the Hausman test is rejected, they can then specify smaller sufficient sets for those individuals with p-values below 10%, re-estimate the model, take these estimates as the reference points for another round of Hausman tests, and so on until the rate of non-rejection is deemed satisfactory.

³² We excluded the results for a small number of women for whom there is not sufficient price variation in their sufficient sets to identify all of the price coefficients, and a small number for whom the estimated price sensitivity in all specifications was positive. Including them in the analysis would change none of the qualitative conclusions drawn from this illustration.

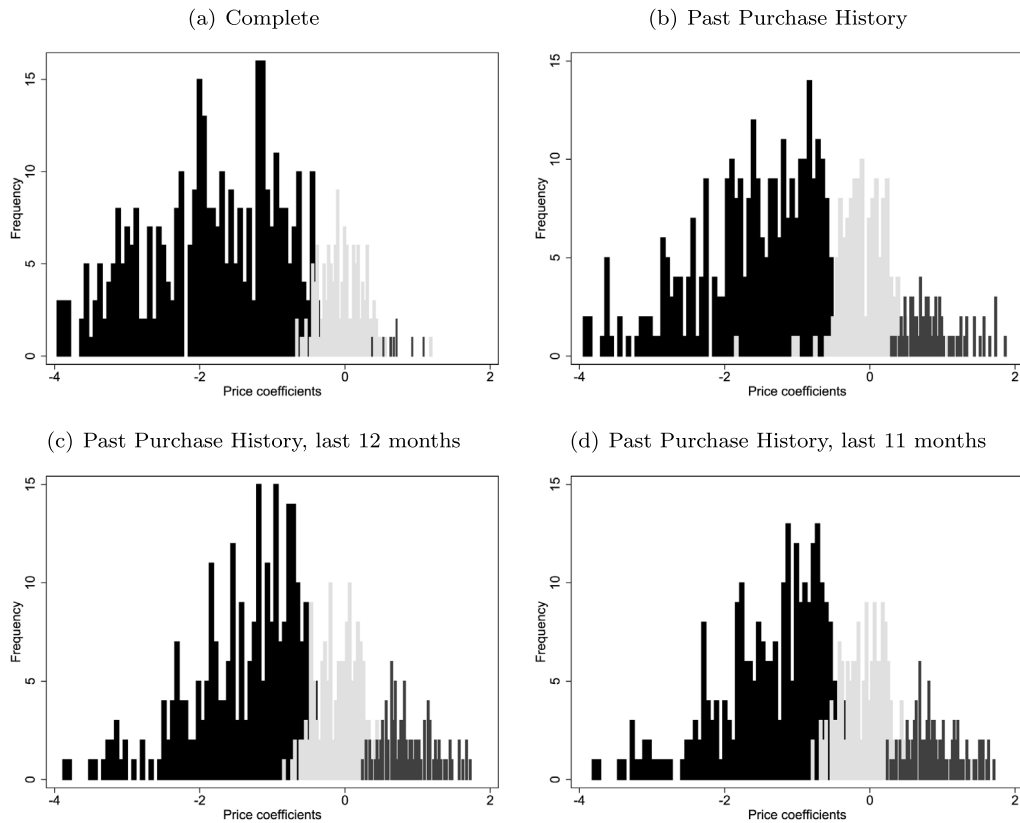


Fig. 2. Distributions of estimated price coefficients. Note: The histograms show the distribution of estimated price coefficients for our sample of adult women for specifications using different sufficient sets. Bars in black are statistically significant and negative, in light gray are statistically not different from zero, and dark gray statistically significant and positive at conventional levels.

Table 4
Hausman tests.

	% of sample with p-value on individual Hausman test			
	>0.1	0.05–0.1	0.01–0.05	<0.01
Price coefficients				
Complete v PPH	12.0	1.9	3.4	82.7
PPH v PPH 12 months	16.2	2.8	5.6	75.4
PPH 12 months v PPH 11 months	37.0	8.5	13.0	41.5
Advertising coefficients				
Complete v PPH	0.0	0.0	0.0	100.0
PPH v PPH 12 months	0.0	8.5	0.0	91.5
PPH 12 months v PPH 11 months	63.0	10.2	0.0	26.9

Notes: The Table summarizes p-values of a Hausman test for each individual.

Our empirical results are in line with Goeree (2008)'s. With respect to price sensitivity, in a simplified model with three products, Goeree (2008, Appendix B, pages 2–7) shows analytically that the more likely are individuals to select among less than the full choice set (what she calls “limited information”), the more attenuated will price elasticities be (i.e., closer to zero). This is also what she finds in her empirical results (Goeree, 2008, Table VII), with price elasticities smaller in absolute value than their full-information counterparts (estimated on what we would call the Complete sufficient set).

Across specifications, we find that our estimates of advertising sensitivity are smaller when using the Past Purchase History sufficient set. As described above, the literature analyzing the economics of advertising has argued that advertising can both inform individuals about products' existence and so increase the likelihood that they are in individuals' choice sets, as well as directly influence individual utility, shifting their preferences. The estimates using the Complete sufficient

Table 5
Complementary value of advertising.

	Complete	PPH	PPH 12 months	PPH 11 months
Mean	0.756	0.656	0.403	0.404
Std Dev	0.168	0.236	0.249	0.261

Notes: The Table reports the mean and standard deviation of the estimated complementary value of advertising. Each column shows the estimates for the indicated sufficient set.

set can, at some intuitive level, be considered as a “reduced form” that captures both of these effects, while the estimates using the Past Purchase History sufficient sets, by focusing on those products for which individual attention is presumed to be already high, identify the effects of advertising mainly through its influence on preferences. If this story is an accurate characterization of behavior in the on-the-go chocolate market, then we would expect to find, as we do, *smaller* estimated advertising sensitivity with the Past Purchase History than with the Complete sufficient set.

In [Becker and Murphy \(1993\)](#), advertising enters consumers’ utility functions as a complement to the value they place on a good being advertised. As such, it has a value to consumers that can be quantified in a manner similar to any other product characteristic. Such a specification is common in the empirical analysis of the effects of advertising, and is something in which firms and advertising executives are likely to be interested. We can use the estimated preference parameters and Eq. (4.1) to compute the complementary value (to consumers) of advertising; these are in [Table 5](#).

These estimates show that different assumptions about sufficient sets may have important practical consequences and lead to very different economic implications. At mean advertising, a one-standard deviation increase in the log advertising stock, $\ln(a_{ibt})$, equal to 0.69 (or 69%), implies an increase in valuation of a product of 52.2 pence when using the Complete sufficient set.³³ As the average price of a chocolate product is 58 pence, this is a 90% increase. By contrast, the estimates obtained using the PPH 12 months sufficient set suggest a one-standard deviation increase in the log advertising stock increases the value of a product by 27.8 pence, or a 48% increase.³⁴

6. Conclusion

In this paper, we survey the two main empirical approaches to tackling the problem of unobserved choice set heterogeneity: “integrating over” and “differencing out” unobserved choice sets. The two approaches originate from different econometric literatures, started respectively by [Manski \(1977\)](#) and [McFadden \(1978\)](#). While integrating over heterogeneous unobserved choice sets is commonly done in empirical applications, differencing them out appears to be less popular in this context, possibly because the [McFadden \(1978\)](#)’s original motivation was to facilitate estimation with large but *observed* choice sets. We provide a unifying notation for understanding the two approaches and, inspired by [Chamberlain \(1980\)](#), we propose the use of consumers’ observed choices paired with assumptions about the evolution of their unobserved choice sets over time as a practical tool to construct proper choice *subsets* in panel data environments. We call these subsets “sufficient sets”.

Sufficient sets serve several purposes. First, sufficient sets help clarify that differencing out can also address the problem of unobserved choice sets, and that it is complementary to integrating over them. Second, sufficient sets prove useful to implement both approaches in practice. Third, they help translate economic assumptions derived from the characteristics of a given choice environment into econometric assumptions appropriate for estimation.

We illustrate some of the relevant issues and methods both in Monte Carlo simulations and in an empirical illustration of on-the-go demand for chocolate bars in the UK. Both exercises highlight how different assumptions on individuals’ choice sets will have a material impact on demand estimation and that special care needs to be taken in analyzing data subject to unobserved choice set heterogeneity.

Appendix A. Importance sampling procedure from [Goeree \(2008\)](#)

In this Appendix, we detail how to implement the importance sampling procedure proposed by [Goeree \(2008\)](#) for the estimation of model (3.2) when choice sets are potentially large. To obtain further computational simplifications, this procedure can be combined with the use of sufficient sets, as described in Section 3.3.

The unconditional probability of individual i choosing alternative j_t in choice situation t is:

³³ $(0.69 \times 0.756) = 0.522$, where 0.756 is the mean complementary value of advertising from [Table 5](#) using the Complete sufficient set.

³⁴ $(0.69 \times 0.403) = 0.278$, where 0.403 is the mean complementary value of advertising from [Table 5](#) using the Past Purchase History for the past 12 months.

$$\Pr[Y_{it} = j_t | \theta, \gamma] = \sum_{c_t \in C_t^j} \frac{\exp(V(X_{ijt}, \theta))}{\sum_{d_t \in c_t} \exp(V(X_{idt}, \theta))} \times \overbrace{\prod_{l_t \in c_t} \phi_{ilt}(\gamma) \prod_{k_t \notin c_t} (1 - \phi_{ikt}(\gamma))}^{\Pr[CS_{it}^* = c_t | \gamma]}, \quad (\text{A.1})$$

where C_t^j is the collection of all possible choice sets that include alternative j_t in period t , $\Pr[Y_{it} = j_t | CS_{it}^* = c_t, \theta]$ is the choice probability conditional on choice set $CS_{it}^* = c_t$, and $\Pr[CS_{it}^* = c_t | \gamma]$ is the probability of i being matched to choice set c_t at choice situation t . Individual i 's probability of alternative l_t to be in their choice set in t is:

$$\phi_{ilt}(\gamma) = \frac{\exp(W_{ilt}\gamma)}{1 + \exp(W_{ilt}\gamma)}, \quad (\text{A.2})$$

where γ are the choice set generating process parameters. The aim is to estimate both θ and γ by maximum likelihood on the basis of (A.1). Doing this directly is often numerically infeasible because the set of possible choice sets in each t is usually too large to handle. Goeree (2008) proposed a simulation method to ease the computation of (A.1). In the basic version of it, for each i and t one would approximate (A.1) by drawing R choice sets $\{c_{it}^r | r = 1, \dots, R\}$ according to probability $\Pr[CS_{it}^* = c_{it}^r | \gamma]$ and then by averaging out across the resulting conditional choice probabilities:³⁵

$$\widehat{\Pr}[Y_{it} = j_t | \theta, \gamma] = R^{-1} \sum_{r=1}^R \Pr[Y_{it} = j_t | CS_{it}^* = c_{it}^r, \theta]. \quad (\text{A.3})$$

An estimator based on (A.3) would still be numerically demanding since the probability with which each c_{it}^r is drawn, $\Pr[CS_{it}^* = c_{it}^r | \gamma]$, is a function of γ . This means that at each iteration of the maximization routine, one would have to re-draw the R simulated choice sets $\{c_{it}^r | r = 1, \dots, R\}$ for each i and t . To overcome also this problem, Goeree (2008) proposed an importance sampling version of simulator (A.3) that allows her to draw all the choice sets once and for all at the beginning of estimation. The idea of the importance sampling is that we wish to draw random variable c from probability $p(c)$ but we are not able to directly. However, we know how to draw from probability $g(c)$ and we have a closed-form solution for expression $\frac{p(c)}{g(c)}$. Consequently, one can draw several c^r 's from $g(c^r)$, multiply each draw c^r by $\frac{p(c^r)}{g(c^r)}$, and the resulting distribution of the drawn c^r 's will be the desired $p(c^r)$. In our context, the desired probability is $p(c_{it}^r) = \prod_{l_t \in c_{it}^r} \phi_{ilt}(\gamma) \prod_{k_t \notin c_{it}^r} (1 - \phi_{ikt}(\gamma))$, while $g(c_{it}^r) = \prod_{l_t \in c_{it}^r} \phi_{ilt}^0 \prod_{k_t \notin c_{it}^r} (1 - \phi_{ikt}^0)$ where ϕ_{ilt}^0 is (A.2) evaluated at some initial guess γ_0 that will be picked at the beginning of estimation and will not change until the end. Then, the importance sampling simulator of (A.1) is:

$$\begin{aligned} \widehat{\Pr}[Y_{it} = j_t | \theta, \gamma] &= R^{-1} \sum_{r=1}^R \frac{p(c_{it}^r)}{g(c_{it}^r)} \times \Pr[Y_{it} = j_t | CS_{it}^* = c_{it}^r, \theta] \\ &= R^{-1} \sum_{r=1}^R \left[\frac{\prod_{l_t \in c_{it}^r} \phi_{ilt}(\gamma) \prod_{k_t \notin c_{it}^r} (1 - \phi_{ikt}(\gamma))}{\prod_{l_t \in c_{it}^r} \phi_{ilt}^0 \prod_{k_t \notin c_{it}^r} (1 - \phi_{ikt}^0)} \times \Pr[Y_{it} = j_t | CS_{it}^* = c_{it}^r, \theta] \right]. \end{aligned} \quad (\text{A.4})$$

Simulator (A.4) can be implemented with the following algorithm. Before starting estimation, one should draw R choice sets $\{c_{it}^r | r = 1, \dots, R\}$ and compute their "drawing" probabilities $g(c_{it}^r)$ for each i and t . This can be done as follows.

1. Set some initial value for γ and call it γ_0 .
2. Given γ_0 , compute $\phi_{ilt}^0 = \phi_{ilt}(\gamma_0) = \frac{\exp(W_{ilt}\gamma_0)}{1 + \exp(W_{ilt}\gamma_0)}$ for each i , alternative $l_t \in J_t$, and t .³⁶
3. For each i and t , draw R choice sets $\{c_{it}^r | r = 1, \dots, R\}$ from $\phi_{ilt}^0, l \in J_t$. Each choice set c_{it}^r can be drawn as follows:
 - (a) Draw an independent uniform $u_{ilt}^r \in [0, 1]$ for each $l_t \in J_t$.
 - (b) Alternative $l_t \in J_t$ belongs to c_{it}^r if and only if $u_{ilt}^r \leq \phi_{ilt}^0$.

³⁵ As a useful complement, Honka (2014) proposed a kernel smoothing procedure to prevent lumpiness in these simulated choice probabilities even for "manageable" numbers of draws, R .

³⁶ Remember that J_t is the set of alternatives available in the market in t .

4. For each i and t , compute the probability of having drawn each of the R choice sets $\{c_{it}^r \mid r = 1, \dots, R\}$ as $g(c_{it}^r) = \prod_{l \in c_{it}^r} \phi_{il_t}^0 \prod_{k_t \notin c_{it}^r} (1 - \phi_{ik_t}^0)$.

Then, given $\{c_{it}^r \mid r = 1, \dots, R\}$ and their “drawing” probabilities $g(c_{it}^r)$ for each i and t , one can proceed to the estimation of θ and γ by simulated maximum likelihood.

1. For each guessed value of (θ, γ) , compute the predicted probability of the observed choice sequence of each i as:

$$\widehat{\Pr}[Y_i = (Y_{i1} = j_1, \dots, Y_{it} = j_t, \dots, Y_{iT} = j_T) \mid \theta, \gamma] = \prod_{t=1}^T \widehat{\Pr}[Y_{it} = j_t \mid \theta, \gamma], \quad (\text{A.5})$$

where each $\widehat{\Pr}[Y_{it} = j_t \mid \theta, \gamma]$ in the product is computed as in (A.4) given $\{c_{it}^r \mid r = 1, \dots, R\}$ and their probabilities $g(c_{it}^r)$.

2. Take the log of the individual likelihood contribution from (A.5) and sum across all individuals.
3. Keep iterating with new guesses of (θ, γ) until the log-likelihood function computed at the previous step is maximized.

Appendix B. Derivation of sufficient set logit (SSL) model (3.5)

$$\begin{aligned} \Pr[Y_i = j \mid f(Y_i) = r, \theta] &= \Pr[Y_i = j \mid f(Y_i) = r, \mathcal{CS}_i^* = c, \theta] \\ &= \frac{\Pr[Y_i = j, Y_i \in r, \mathcal{CS}_i^* = c \mid \theta, \gamma]}{\Pr[Y_i \in r, \mathcal{CS}_i^* = c \mid \theta, \gamma]} \\ &= \frac{\Pr[Y_i = j, Y_i \in r \mid \mathcal{CS}_i^* = c, \theta] \Pr[\mathcal{CS}_i^* = c \mid \gamma]}{\Pr[Y_i \in r \mid \mathcal{CS}_i^* = c, \theta] \Pr[\mathcal{CS}_i^* = c \mid \gamma]} \\ &= \frac{\Pr[Y_i = j, Y_i \in r \mid \mathcal{CS}_i^* = c, \theta]}{\sum_{k \in \mathcal{U}} \Pr[Y_i = k, Y_i \in r \mid \mathcal{CS}_i^* = c, \theta]} \\ &= \frac{\prod_{t=1}^T \frac{\exp(V(X_{ijt}, \theta))}{\sum_{v_t \in \mathcal{CS}_{it}^* = c_t} \exp(V(X_{iv_t}, \theta))}}{\sum_{k \in f(Y_i)=r} \prod_{t=1}^T \frac{\exp(V(X_{ikt}, \theta))}{\sum_{v_t \in \mathcal{CS}_{it}^* = c_t} \exp(V(X_{iv_t}, \theta))}} \\ &= \frac{\prod_{t=1}^T \exp(V(X_{ijt}, \theta))}{\sum_{k \in f(Y_i)=r} \prod_{t=1}^T \exp(V(X_{ikt}, \theta))} \end{aligned} \quad (\text{B.1})$$

Assumption 1 and **Condition 1** imply the IIA property, and the first equality follows from its definition. Note that conditioning the choice probability on $f(Y_i) = r$ is equivalent to conditioning the choice Y_i to be from the set r , or $Y_i \in r$. The second and third equalities follow from the definition of conditional probability, while the fourth follows from the law of total probability. In the fourth equality, \mathcal{U} is the universal set of *all* choice sequences. The fifth equality follows from $\Pr[Y_i = k, Y_i \in r \mid \mathcal{CS}_i^* = c, \theta]$ being equal to $\Pr[Y_i = k \mid \mathcal{CS}_i^* = c, \theta]$ for any $k \in r$ or, alternatively, being equal to 0 for any $k \notin r$. In the last equality, $\sum_{v_t \in \mathcal{CS}_{it}^* = c_t} \exp(V(X_{iv_t}, \theta))$ cancels out. Finally, consistency of the conditional Maximum Likelihood Estimator derived from $\Pr[Y_i = j \mid f(Y_i) = r, \theta]$ follows from [McFadden \(1978\)](#).

Appendix C. Derivation of sufficient set logit (SSL) model (3.6)

In this Appendix we demonstrate that Eq. (3.6) holds if and only if $f(Y_i) = \times_{t=1}^T f_t(Y_i)$.

IF part. Suppose that $f(Y_i) = \times_{t=1}^T f_t(Y_i)$. Then we can re-write the denominator of conditional logit model (3.5), $\Pr[Y_i = j | f(Y_i) = r, \theta]$, as (omitting the f in the conditioning for simplicity):

$$\begin{aligned} \sum_{(k_1, \dots, k_T) \in r} \prod_{t=1}^T \exp(V_{ik_t t}) &= \sum_{(k_1, \dots, k_T) \in r_1 \times \dots \times r_T} \prod_{t=1}^T \exp(V_{ik_t t}) \\ &= \left(\sum_{k_1 \in r_1} \exp(V_{ik_1 1}) \right) \sum_{(k_2, \dots, k_T) \in r_2 \times \dots \times r_T} \prod_{t=2}^T \exp(V_{ik_t t}) \\ &= \left(\sum_{k_1 \in r_1} \exp(V_{ik_1 1}) \right) \left(\sum_{k_2 \in r_2} \exp(V_{ik_2 2}) \right) \sum_{(k_3, \dots, k_T) \in r_3 \times \dots \times r_T} \prod_{t=3}^T \exp(V_{ik_t t}) \\ &\vdots \\ &= \prod_{t=1}^T \left(\sum_{k_t \in r_t} \exp(V_{ik_t t}) \right), \end{aligned} \quad (C.1)$$

which implies that:

$$\begin{aligned} \Pr[Y_i = j | f(Y_i) = r, \theta] &= \frac{\prod_{t=1}^T \exp(V_{ij_t t})}{\sum_{(k_1, \dots, k_T) \in r} \prod_{t=1}^T \exp(V_{ik_t t})} \\ &= \frac{\prod_{t=1}^T \exp(V_{ij_t t})}{\prod_{t=1}^T \left(\sum_{k_t \in r_t} \exp(V_{ik_t t}) \right)} \\ &= \prod_{t=1}^T \frac{\exp(V_{ij_t t})}{\sum_{k_t \in r_t} \exp(V_{ik_t t})} \\ &= \prod_{t=1}^T \Pr[Y_{it} = j_t | f_t(Y_i) = r_t, \theta]. \end{aligned} \quad (C.2)$$

To complete the proof, we are now going to show that $\Pr[Y_{it} = j_t | f(Y_i) = r, \theta] = \Pr[Y_{it} = j_t | f_t(Y_i) = r_t, \theta]$. Define the set $M(\tilde{j}_s) = \{(z_1, \dots, z_s, \dots, z_T) | z \in f(Y_i) = r, z_s = \tilde{j}_s\}$ as the collection of choice sequences that have alternative \tilde{j}_s in position s . It then follows that:

$$\begin{aligned} \Pr[Y_{is} = \tilde{j}_s | f(Y_i) = r, \theta] &= \sum_{j \in M(\tilde{j}_s)} \Pr[Y_i = j | f(Y_i) = r, \theta] \\ &= \sum_{j \in M(\tilde{j}_s)} \frac{\prod_{t=1}^T \exp(V_{ij_t t})}{\sum_{(k_1, \dots, k_T) \in r} \prod_{t=1}^T \exp(V_{ik_t t})} \\ &= \left(\sum_{(k_1, \dots, k_T) \in r} \prod_{t=1}^T \exp(V_{ik_t t}) \right)^{-1} \left(\sum_{j \in M(\tilde{j}_s)} \prod_{t=1}^T \exp(V_{ij_t t}) \right). \end{aligned} \quad (C.3)$$

Similarly to (C.1), $f(Y_i) = \times_{t=1}^T f_t(Y_i)$ implies that $M(\tilde{j}_s) = r_1 \times \dots \times \{\tilde{j}_s\} \times \dots \times r_T$, and consequently that the numerator of (C.3) can be re-written as:

$$\begin{aligned} \sum_{j \in M(\tilde{j}_s)} \prod_{t=1}^T \exp(V_{ijt}) &= \sum_{(j_1, \dots, j_s, \dots, j_T) \in r_1 \times \dots \times \{\tilde{j}_s\} \times \dots \times r_T} \prod_{t=1}^T \exp(V_{ijt}) \\ &= \exp(V_{\tilde{j}_s s}) \prod_{t \neq s} \left(\sum_{j_t \in r_t} \exp(V_{ijt}) \right). \end{aligned} \quad (C.4)$$

Plugging (C.1) and (C.4) into (C.3), we obtain:

$$\begin{aligned} &\Pr[Y_{is} = \tilde{j}_s | f(Y_i) = r, \theta] \\ &= \sum_{j \in M(\tilde{j}_s)} \Pr[Y_i = j | f(Y_i) = r, \theta] \\ &= \left(\prod_{t=1}^T \left(\sum_{k_t \in r_t} \exp(V_{ikt}) \right) \right)^{-1} \left(\exp(V_{\tilde{j}_s s}) \prod_{t \neq s} \left(\sum_{j_t \in r_t} \exp(V_{ijt}) \right) \right) \\ &= \left(\left(\sum_{k_s \in r_s} \exp(V_{iks}) \right) \prod_{t \neq s} \left(\sum_{k_t \in r_t} \exp(V_{ikt}) \right) \right)^{-1} \left(\exp(V_{\tilde{j}_s s}) \prod_{t \neq s} \left(\sum_{j_t \in r_t} \exp(V_{ijt}) \right) \right) \\ &= \frac{\exp(V_{\tilde{j}_s s})}{\sum_{k_s \in r_s} \exp(V_{iks})} \\ &= \Pr[Y_{is} = \tilde{j}_s | f_s(Y_i) = r_s, \theta]. \end{aligned} \quad (C.5)$$

ONLY IF part. Consider two choice situations t and s . For these, define $f_t = \{j_t | (j_1, \dots, j_t, \dots, j_T) \in f(Y_i) = r\}$ and $f_s = \{j_s | (j_1, \dots, j_s, \dots, j_T) \in f(Y_i) = r\}$ as the collections of alternatives that appear in at least one sequence belonging to $f(Y_i) = r$ at positions t and s , respectively. Suppose $f(Y_i) \neq \times_{t=1}^T f_t(Y_i)$, then $\exists t$ and s such that $(\tilde{j}_t \in f_t, \tilde{j}_s \in f_s)$ and $(\tilde{j}_1, \dots, \tilde{j}_t, \dots, \tilde{j}_s, \dots, \tilde{j}_T) \notin f(Y_i) = r$. It then follows that:

$$\Pr[Y_{it} = \tilde{j}_t | Y_{is} = \tilde{j}_s, f(Y_i) = r, \theta] = 0,$$

while, since $\tilde{j}_t \in f_t$:

$$\Pr[Y_{it} = \tilde{j}_t | f(Y_i) = r, \theta] > 0.$$

This implies that Y_{it} and Y_{is} are not conditionally independent.

Appendix D. Monte Carlo evidence on the performance of SSL model

In this Appendix, we report the results of Monte Carlo experiments evaluating the practical performance of MNL and SSL models in the presence of various forms of unobserved choice set heterogeneity. In Table 6, we directly vary the extent of choice set heterogeneity by randomly removing alternatives from choice sets, independently of the indirect utilities or the product characteristics of the removed alternatives. Differently, in Table 7 we implement two more economically relevant choice set formation processes: a model of screening on product characteristics (such as price) and a model of costly search. Here, our aim is to illustrate that even when choice set heterogeneity is the outcome of selection processes involving the alternatives' systematic utilities and/or product characteristics, the proposed SSL models work well without requiring the econometrician to know much about such possibly complex processes.

The first column of Table 6 reports results showing the bias in a MNL model from incorrectly assuming that all individuals in all choice situations have access to the full choice set, made of five alternatives. The second column reports estimates of the true MNL model, i.e. the model that correctly assigns the true choice set facing each individual in each choice situation. There is no estimation bias in this case. The other columns report estimates from, respectively, the Full Purchase History (FPH), the Past Purchase History (PPH), and the Choice Permutation (CP) SSL models.

Table 6
Performance of sufficient set logits.

	MNL, full % Bias	MNL, true % Bias	FPH SSL % Bias	PPH SSL % Bias	CP SSL % Bias
Baseline					
100% full choice set	0.3%	0.3%	0.6%	0.8%	1.2%
Increasing share of individuals with a random product removed from choice set					
10% constrained	11.2%	0.2%	0.4%	0.9%	0.8%
30% constrained	26.3%	0.3%	0.6%	1.2%	0.6%
50% constrained	36.0%	0.4%	0.7%	1.0%	0.9%
Increasing share of products randomly removed from choice set					
30% have 4 of 5	26.3%	0.3%	0.6%	1.2%	0.6%
30% have 3 of 5	36.0%	0.7%	1.0%	1.3%	1.2%
30% have 2 of 5	57.0%	0.4%	0.5%	0.7%	0.5%

We consider a population of 1000 individuals making a sequence of choices over 10 choice situations. On each choice situation, they choose between a maximum of five alternatives. The indirect utility of each alternative is specified as in Eq. (2.1). The systematic utility is $V(X_{ijt}, \theta) = \delta_{jt} + X_{ijt}\beta$, and the unobserved portion of utility, ϵ_{ijt} , is distributed i.i.d. Gumbel. X_{ijt} is drawn from a normal distribution with mean 0 and variance 5, $\delta_{jt} = 0$ for all j 's, and $\beta = 2$. The table reports averages of the percentage absolute bias of the estimates, $|(\hat{\beta} - \beta)/\beta| \times 100$. In the top panel, all individuals make choices from the full choice set. In the central panel, an increasing share of individuals make choices from a choice set of four randomly selected alternatives. In the bottom panel, 30% of individuals make choices from a choice set of two, three, or four randomly selected alternatives. In the central and bottom panels, choice sets differ across individuals but are constant across choice situations within individual. We simulate and average results over 20 replications per scenario. To speed up computations, the CP SSL is estimated by sampling at random (uniformly), for each individual, 5000 permutations of the observed sequence of choices, as suggested by D'Haultfoeulle and Iaria (2016).

The top panel of Table 6 shows the lack of bias in the absence of unobserved choice set heterogeneity. The following two panels show, in turn, the bias arising from, first, increasing the share of individuals with restricted choice sets and, second, increasing the severity of the restriction on choice sets. Overall, there is significant bias when we incorrectly assume full choice sets (the first column), but that there is no average bias when relying on any of these three SSL models for estimation.

Table 7 reports results for two economically relevant choice set formation processes: a model of screening on product characteristics in the central panel and a model of costly search in the bottom panel. Both the models of screening and of costly search are simple. In these simulations, our aim is not to implement the most realistic screening and search models that have appeared in the literature, but rather to study the performance of MNL and SSL models when choice set heterogeneity is the outcome of non-trivial selection processes involving the alternatives' systematic utilities and/or product characteristics.

The first column of Table 7 reports results for a MNL with a full choice set of five alternatives. The second column reports results for the MNL with true choice sets, as if one could perfectly observe the outcomes of the screening and costly search for each individual in each choice situation. Both models of screening and of costly search generate choice sets that are weakly growing over choice situations, compatibly with the assumptions of the PPH sufficient set. The third column reports the estimates of a PPH SSL model.

The central panel of Table 7 reports results for a choice set formation model of screening on product characteristic X_{ijt} . Each individual i has a maximum threshold \bar{X}_i for the value of X_{ijt} they are willing to consider.³⁷ In $t = 1$, CS_{i1}^* contains those alternatives for which $X_{ij1} \leq \bar{X}_i$.³⁸ Denote by \overline{CS}_{it}^* the collection of alternatives not in CS_{it}^* . Once an alternative is considered in t , it will also be in $CS_{it'}^*$ for $t' > t$. Accordingly, in any $t > 1$, individual i checks whether any of the alternatives in $\overline{CS}_{i,t-1}^*$, i.e. those *not* already in $CS_{i,t-1}^*$, has an acceptable value of X_{ijt} and includes in CS_{it}^* all those for which $X_{ijt} \leq \bar{X}_i$. In other words, in each $t > 1$, CS_{it}^* is the union between $CS_{i,t-1}^*$ and those alternatives from $\overline{CS}_{i,t-1}^*$ that pass the \bar{X}_i screening.

The bottom panel of Table 7 reports results for a choice set formation model of costly search over alternatives. Each individual i in every t , given the set of alternatives already in their choice set from $t-1$, $CS_{i,t-1}^*$, considers whether to incur a search cost of c_{ij} to include any *new* alternative in CS_{it}^* (i.e., any alternative belonging to $\overline{CS}_{i,t-1}^*$). When considering whether to add or not an additional alternative to the choice set, individuals perfectly observe all the X_{ijt} 's and search costs, but need to form expectations about the ϵ_{ijt} error terms (according to Assumption 1, the choice set formation process cannot

³⁷ Each X_{ijt} is distributed normal with mean 0 and variance 5. The individual-specific threshold \bar{X}_i is distributed standard normal.

³⁸ To prevent CS_{i1}^* from being empty, a randomly selected alternative is included in CS_{i1}^* when $X_{ij1} > \bar{X}_i$ for all alternatives.

Table 7
A Model of screening and a model of search.

	MNL, full % Bias	MNL, true % Bias	PPH SSL % Bias
Baseline			
100% full choice set	0.3%	0.3%	0.8%
Increasing share of individuals who screen sequentially			
30% screens	39.5%	0.2%	0.4%
50% screens	50.3%	0.1%	0.1%
90% screens	63.5%	0.1%	0.4%
Increasing share of individuals who search sequentially			
30% searches	52.8%	0.5%	1.6%
50% searches	64.3%	0.2%	1.6%
90% searches	77.3%	0.3%	2.6%

We consider a population of 1000 individuals making a sequence of choices over 10 choice situations. On each choice situation, they choose between a maximum of five alternatives. The indirect utility of each alternative is specified as in Eq. (2.1). The systematic utility is $V(X_{ijt}, \theta) = \delta_{jt} + X_{ijt}\beta$, and the unobserved portion of utility, ϵ_{ijt} , is distributed i.i.d. Gumbel. X_{ijt} is drawn from a normal distribution with mean 0 and variance 5, $\delta_{jt} = 0$ for all j_t 's, and $\beta = 2$. The table reports averages of the percentage absolute bias of the estimates, $|\hat{\beta} - \beta|/\beta \times 100$. In the top panel, all individuals make choices from the full choice set. In the central panel, an increasing share of individuals make choices from a choice set formed sequentially by screening over the X_{ijt} 's (see text for detail). In the bottom panel, an increasing share of individuals make choices from a choice set formed sequentially by searching over alternatives (see text for detail). In the central and bottom panels, choice sets differ across individuals and evolve across choice situations within individual. We simulate and average results over 20 replications per scenario.

depend on the realizations of the error terms).³⁹ In $t = 1$, CS_{i1}^* contains those alternatives for which $V(X_{ij1}, \theta) - c_{ij} \geq 0$.⁴⁰ In any $t > 1$, individual i is assumed to be able to add to their choice set at most one alternative from $\overline{CS}_{i,t-1}^*$ (i.e., either add one alternative or nothing). Similar to the model of screening, once an alternative is considered in t , it will also be in CS_{it}^* for $t' > t$. In any $t > 1$, individual i decides to search for an additional alternative to be included in CS_{it}^* only when the expected net benefit from searching is greater than the expected maximal utility from $CS_{i,t-1}^*$ (i.e., what can be achieved without any additional search):

$$\max_{j_t \in \overline{CS}_{i,t-1}^*} \{V(X_{ijt}, \theta) - c_{ij}\} \geq \ln \left[\sum_{k_t \in CS_{i,t-1}^*} \exp(V(X_{ikt}, \theta)) \right], \quad (D.1)$$

where $\overline{CS}_{i,t-1}^*$ is the collection of alternatives not included in $CS_{i,t-1}^*$, the choice set including all the alternatives searched for in the previous choice situations. When i searches in t , the alternative in $\overline{CS}_{i,t-1}^*$ corresponding to the largest expected net benefit from searching is included in CS_{it}^* .

Overall, Table 7 shows that when the choice set formation process is a function of the alternatives' systematic utilities and/or product characteristics, mistakenly ignoring it may have detrimental effects on the estimation of preference parameters (first column). Clearly, if one had data on the true choice sets faced by each individual in each choice situation, then neither choice set generating process would cause any estimation problem given that Assumption 1 still holds (second column). Finally, the third column shows that the PPH SSL performs virtually as well as the true MNL (second column), with the advantage of not requiring the econometrician to have any additional data on true choice sets or to know much about the potentially complex details of the choice set generating process.

Appendix E. Confidence intervals for elasticity bounds

As an example of how to conduct inference on the identification regions described in Section 4.2, we construct confidence intervals for the elasticity bounds following Imbens and Manski (2004). For notational simplicity, we limit our discussion to a single elasticity term $\xi_{it}^{jk}(X_{it}, \theta)$, although the same ideas can be extended to the collection of all elasticities.

³⁹ Each X_{ijt} is distributed normal with mean 0 and variance 5. The individual-alternative specific search cost c_{ij} is distributed log-normal with mean 3 and variance 1. Each ϵ_{ijt} error term is distributed Gumbel. Individuals have correct beliefs about the distribution of the error terms when computing expected utilities.

⁴⁰ When computing expected utilities, we ignore the Euler constant. This is an approximation only in $t = 1$, for any $t > 1$ the constant does indeed drop out of rule (D.1). To prevent CS_{i1}^* from being empty, a randomly selected alternative is included in CS_{i1}^* when $V(X_{ij1}, \theta) - c_{ij} < 0$ for all alternatives.

Refer to the upper and lower bounds of $\xi_{it}^{jk}(X_{it}, \theta)$ in (4.9) as to $\overline{\xi_{it}^{jk}}(X_{it}, \theta)$ and $\underline{\xi_{it}^{jk}}(X_{it}, \theta)$, respectively. Denote the elasticity bounds of $\xi_{it}^{jk}(X_{it}, \theta)$ by the 2×1 vector $B(\xi_{it}^{jk}(X_{it}, \theta)) = [\underline{\xi_{it}^{jk}}(X_{it}, \theta), \overline{\xi_{it}^{jk}}(X_{it}, \theta)]'$ and the corresponding elasticity interval from (4.9) by $IN(\xi_{it}^{jk}(X_{it}, \theta))$. Then, given X_{it} and our consistent $\hat{\theta}$, we can estimate the elasticity bounds $B(\xi_{it}^{jk}(X_{it}, \theta))$ by $B(\xi_{it}^{jk}(X_{it}, \hat{\theta}))$. We derive the corresponding $100(1 - \alpha)$ percent confidence interval $CI_{1-\alpha}$ from condition:

$$\inf_{\xi_{it}^{jk} \in IN(\xi_{it}^{jk}(X_{it}, \theta))} \left\{ \lim_{I \rightarrow \infty} \Pr \left[\xi_{it}^{jk} \in CI_{1-\alpha} \right] \right\} \geq 1 - \alpha. \quad (E.1)$$

Since our estimator is consistent and asymptotically normal, i.e., $\hat{\theta}\sqrt{I} \xrightarrow{d} \mathcal{N}(\theta, V_\theta)$, by the delta-method:

$$B(\xi_{it}^{jk}(X_{it}, \hat{\theta}))\sqrt{I} \xrightarrow{d} \mathcal{N} \left(B(\xi_{it}^{jk}(X_{it}, \theta)), \frac{\partial B(\xi_{it}^{jk}(X_{it}, \theta))}{\partial \theta'} V_\theta \frac{\partial B(\xi_{it}^{jk}(X_{it}, \theta))'}{\partial \theta'} \right). \quad (E.2)$$

Refer to the 2×2 asymptotic variance-covariance matrix of $B(\xi_{it}^{jk}(X_{it}, \hat{\theta}))$ as to $\Sigma_{B(\xi_{it}^{jk})}$. It follows that, whenever $f_t(Y_i) \cup \{j_t\} = r_{it} \cup \{j_t\}$ is a strict subset of $Q_{it} \cup \{j_t\} = q_{it} \cup \{j_t\}$, so that for any X_{it} and θ , $\underline{\xi_{it}^{jk}}(X_{it}, \theta) < \overline{\xi_{it}^{jk}}(X_{it}, \theta)$, condition (E.1) is satisfied by:

$$CI_{1-\alpha} = \left[\underline{\xi_{it}^{jk}}(X_{it}, \hat{\theta}) - q_{1-\alpha} \sqrt{\Sigma_{B(\xi_{it}^{jk})}^{11}}, \overline{\xi_{it}^{jk}}(X_{it}, \hat{\theta}) + q_{1-\alpha} \sqrt{\Sigma_{B(\xi_{it}^{jk})}^{22}} \right], \quad (E.3)$$

where $q_{1-\alpha}$ is the $(1 - \alpha)^{th}$ quantile of the standard normal distribution.

In the extreme case in which $f_t(Y_i) \cup \{j_t\} = r_{it} \cup \{j_t\} = Q_{it} \cup \{j_t\} = q_{it} \cup \{j_t\}$, $\underline{\xi_{it}^{jk}}(X_{it}, \theta) = \overline{\xi_{it}^{jk}}(X_{it}, \theta)$ for any X_{it} and θ , and (E.3) is invalid. This is due to a discontinuity at $\xi_{it}^{jk}(X_{it}, \theta) = \overline{\xi_{it}^{jk}}(X_{it}, \theta)$, since in that case the coverage of the interval is only $100(1 - 2\alpha)\%$ rather than the nominal $100(1 - \alpha)\%$. (See [Imbens and Manski, 2004](#) for a modification of (E.3) that overcomes this problem.) However, note that (a) both $f_t(Y_i) \cup \{j_t\} = r_{it} \cup \{j_t\}$ and $Q_{it} \cup \{j_t\} = q_{it} \cup \{j_t\}$ are always perfectly observed by the econometrician, so that the appropriate $CI_{1-\alpha}$ can always be implemented and that (b) in our empirical application $f_t(Y_i) \cup \{j_t\} \subset Q_{it} \cup \{j_t\}$ for every i and t .

Appendix F. Specification tests: Choosing among sufficient sets

In this Appendix, we first describe how [Ruud \(1984\)](#)'s *Factorization Theorem* can be used to construct specification tests (in the spirit of [Hausman and McFadden, 1984](#)) for SSL and ISSL models that are helpful to discriminate among different sufficient sets. Second, we illustrate with some examples how to use these statistics to test for features of the choice set formation process and of unobserved preference heterogeneity. To keep notation simple, in what follows we focus on the SSL model with the understanding that the same results apply almost verbatim to the ISSL model.

Suppose that [Assumption 1](#) holds, and that sufficient sets f_L and f_Z satisfy [Condition 1](#), that $f_Z(Y_i) \subset f_L(Y_i)$, $Y_i \in \mathcal{CS}_i^* = c$, and that $i = 1, \dots, I$. Define $l_L(\theta)$ and $l_Z(\theta)$ as the log-likelihood functions corresponding to the SSL models with sufficient sets $f_L(Y_i)$ and $f_Z(Y_i)$, and denote by $\hat{\theta}_L$ and $\hat{\theta}_Z$ the corresponding MLEs. Then the following results hold:

1. The log-likelihood function $l_L(\theta)$ can be written as $l_L(\theta) = l_Z(\theta) + l_\Delta(\theta)$.
2. Provided that θ is identified in $l_\Delta(\theta)$, so that $\hat{\theta}_\Delta$ is a well defined MLE, then:

- (a) $\hat{\theta}_Z$ and $\hat{\theta}_\Delta$ are asymptotically independent, and
- (b) $\hat{\theta}_L$ is more efficient than $\hat{\theta}_Z$.

3. Given result (2), then:

- (a) All Hausman tests based on pairwise estimator comparisons among $\hat{\theta}_L$, $\hat{\theta}_Z$, and $\hat{\theta}_\Delta$ are equivalent,
- (b) The Likelihood Ratio statistic $LR = 2[l_Z(\hat{\theta}_Z) + l_\Delta(\hat{\theta}_\Delta) - l_L(\hat{\theta}_L)]$ is asymptotically equivalent to the Hausman statistic comparing $\hat{\theta}_L$ and $\hat{\theta}_Z$, and
- (c) $\text{Var}(\hat{\theta}_L - \hat{\theta}_Z) = \text{Var}(\hat{\theta}_Z) - \text{Var}(\hat{\theta}_\Delta)$.

Proof of result (1). Starting from (3.5), we can re-write for every i the probability of the observed choice sequence j given $f_Z(Y_i) = z \subset f_L(Y_i) = l$ as:

$$\begin{aligned} \Pr[Y_i = j | f_L(Y_i) = l, \theta] &= \frac{\prod_{t=1}^T \exp(V(X_{ijt}, \theta))}{\sum_{k \in f_L(Y_i)=l} \prod_{t=1}^T \exp(V(X_{ikt}, \theta))} \\ &= \Pr[Y_i = j | f_Z(Y_i) = z, \theta] \left(\frac{\Pr[Y_i = j | f_L(Y_i) = l, \theta]}{\Pr[Y_i = j | f_Z(Y_i) = z, \theta]} \right) \\ &= \frac{\prod_{t=1}^T \exp(V(X_{ijt}, \theta))}{\sum_{q \in f_Z(Y_i)=z} \prod_{t=1}^T \exp(V(X_{iqt}, \theta))} \frac{\sum_{q \in f_Z(Y_i)=z} \prod_{t=1}^T \exp(V(X_{iqt}, \theta))}{\sum_{k \in f_L(Y_i)=l} \prod_{t=1}^T \exp(V(X_{ikt}, \theta))} \\ &= \Pr[Y_i = j | f_Z(Y_i) = z, \theta] \Pr[Y_i \in f_Z(Y_i) = z | f_L(Y_i) = l, \theta], \end{aligned}$$

where $\Pr[Y_i \in f_Z(Y_i) = z | f_L(Y_i) = l, \theta]$ is the probability that a choice sequence belongs to the “smaller” set z relative to the “larger” set l . By multiplying $\Pr[Y_i = j | f_L(Y_i) = l, \theta]$ across all individuals and by taking the logarithm, result (1) follows with $l_\Delta(\theta) = \sum_{i=1}^I \ln(\Pr[Y_i \in f_Z(Y_i) = z | f_L(Y_i) = l, \theta])$.

Proof of result (2). Given result (1), results (2a) and (2b) follow from the *Factorization Theorem* of Ruud (1984, result (1), p.24).

Proof of result (3). Given result (1), result (3a) follows from the *Factorization Theorem* of Ruud (1984, result (3), p.24), while result (3b) follows from (Ruud, 1984, pp. 28–9). Result (3c) can be proved as follows. Ruud (1984, result 2, p.24) shows that $\hat{\theta}_L$ is asymptotically equivalent to $\text{Var}(\hat{\theta}_L) \text{Var}(\hat{\theta}_Z)^{-1} \hat{\theta}_Z + \text{Var}(\hat{\theta}_L) \text{Var}(\hat{\theta}_\Delta)^{-1} \hat{\theta}_\Delta$. This implies that $\text{Cov}(\hat{\theta}_L, \hat{\theta}_Z) = \text{Cov}(\text{Var}(\hat{\theta}_L) \text{Var}(\hat{\theta}_Z)^{-1} \hat{\theta}_Z, \hat{\theta}_Z) = \text{Var}(\hat{\theta}_L) \text{Var}(\hat{\theta}_Z)^{-1} \text{Var}(\hat{\theta}_Z) = \text{Var}(\hat{\theta}_L)$, where the first equality follows from result (2a). Consequently, $\text{Var}(\hat{\theta}_L - \hat{\theta}_Z) = \text{Var}(\hat{\theta}_L) + \text{Var}(\hat{\theta}_Z) - 2\text{Cov}(\hat{\theta}_L, \hat{\theta}_Z) = \text{Var}(\hat{\theta}_Z) - \text{Var}(\hat{\theta}_L)$.

The Likelihood Ratio statistic LR from result (3b) allows one to compare different SSL models derived from alternative assumptions on sufficient sets. It consists of the difference between an unrestricted log-likelihood function, $l_Z(\hat{\theta}_Z) + l_\Delta(\hat{\theta}_\Delta)$, and a restricted one, $l_L(\hat{\theta}_L)$.⁴¹ Even though LR requires the computation of a third estimator, $\hat{\theta}_\Delta$, it is simpler to implement than other Hausman statistics based on quadratic forms. For instance, the statistic LR is always non-negative, bypassing the practical inconvenience of some estimated covariance matrices that fail to be positive definite. In contrast to some other Hausman statistics, LR also makes very transparent the computation of the degrees of freedom of the corresponding χ^2 distribution: they equal the number of parameters in $\hat{\theta}_L$. Result (3c) is of practical convenience, it implies that the computation of $\text{Var}(\hat{\theta}_L - \hat{\theta}_Z)$, necessary for classical Hausman statistics, can proceed as in the standard case in which one of the compared estimators is fully efficient under the null hypothesis, even though no such efficiency assumption is required here.

F.1. Practical examples of testing procedures

For simplicity of exposition, we limit our examples to the SSL model with the understanding that similar ideas readily apply also to the ISSL model, for which the IIA is only assumed *within* each individual (but not *across* individuals). In the context of SSL models, the examples of sufficient sets introduced in section 4.1 rely on the following economic assumptions:

- f_{CP} : Choice set stability across T choice situations and possibility of IIA violations in the form of individual-alternative specific fixed effects, δ_{ijt} .
- f_{FPH} : Choice set stability across T choice situations and IIA property.
- f_{PPH} : Choice set evolution in the form of weakly growing choice sets (or, symmetrically, weakly shrinking choice sets) across T choice situations and IIA property.⁴²

There are two possibilities for making comparisons across SSL models based on different sufficient sets f 's, and each presents ways of implicitly testing for some of the maintained economic assumptions embedded in the compared sufficient sets. The first possibility is to compare f_{CP} , f_{FPH} , and f_{PPH} for choice sequences of constant length T . The second possibility is to fix a specific f , say f_{CP} , and to compare choice sequences with some of their *sub-sequences*: for example, the sequence $1, 2, \dots, T^L$ can be split into two mutually exclusive sub-sequences $1, 2, \dots, T^Z$ and $T^Z + 1, \dots, T^L$, and this gives rise to different f_{CP} 's, f_{CP}^Z and f_{CP}^L such that $f_{CP}^Z(Y_i) \subset f_{CP}^L(Y_i)$ for any $Y_i \in \mathcal{CS}_i^* = c$. We now illustrate with some examples each testing possibility in turn.

⁴¹ As developed more fully in Ruud (1984), this form is common to many econometric tests, including incremental over-identifying (or Sargan) tests commonly used to investigate the validity of subsets of instruments (Arellano, 2003, Section 5.4.4).

⁴² Importantly, the IIA requirement follows from the SSL model and it is not intrinsic in the f_{FPH} and f_{PPH} sufficient sets. Neither sufficient set relies on the IIA property *across* individuals when employed in more general models such as the ISSL and the SSML discussed in Section 3.2.2.

F.1.1. Comparisons of different f 's with constant T

For choice sequences of a given length T , $f_{CP}(Y_i) \subseteq f_{FPH}(Y_i)$ and $f_{PPH}(Y_i) \subseteq f_{FPH}(Y_i)$ for any $Y_i \in \mathcal{CS}_i^* = c$. Suppose $Y_i = (1, 3)$. Then $f_{CP}(1, 3) = \mathcal{P}(1, 3) = \{(1, 3), (3, 1)\}$, $f_{FPH}(1, 3) = \{1, 3\} \times \{1, 3\} = \{(1, 1), (3, 3), (1, 3), (3, 1)\}$, and $f_{PPH}(1, 3) = \{1\} \times \{1, 3\} = \{(1, 1), (1, 3)\}$. Note that there is no clear "inclusion" relationship between $f_{CP}(Y_i)$ and $f_{PPH}(Y_i)$. Given the *Factorization Theorem*, the above relationships among sufficient sets lead to two possible classes of tests. The first is about choice set stability and the second about deviations from the IIA property.

Choice set stability (given IIA property). In the context of SSL models, both f_{FPH} and f_{PPH} rely on the IIA property. However, they rely on different assumptions regarding the evolution of choice sets across choice situations: f_{FPH} assumes that unobserved choice sets do not change along the whole choice sequence, while f_{PPH} allows for the entry of new alternatives in the unobserved choice set while comparing choice situation t to $t + 1$. On the one hand, if unobserved choice sets were stable, then both f 's would give rise to consistent estimators $\hat{\theta}_{FPH}$ and $\hat{\theta}_{PPH}$, but result (2b) tells us that $\hat{\theta}_{FPH}$ would be more efficient than $\hat{\theta}_{PPH}$. On the other hand, if unobserved choice sets were growing over choice situations, then only $\hat{\theta}_{PPH}$ would be consistent: f_{FPH} would not satisfy Condition 1, inducing violations of the IIA property as discussed in Section 2.2. It follows that, under the maintained assumption of the IIA property, a test for H_0 : (choice set stability in $1, 2, \dots, T$) is $LR = 2 [l_{PPH}(\hat{\theta}_{PPH}) + l_{\Delta}(\hat{\theta}_{\Delta}) - l_{FPH}(\hat{\theta}_{FPH})]$.

Departures from IIA property (given choice set stability). The sufficient sets f_{FPH} and f_{CP} are both based on the same assumption of unobserved choice set stability in $1, 2, \dots, T$. However, in the context of SSL models, they rely on different assumptions regarding unobserved preference heterogeneity: f_{FPH} relies on the IIA property, while f_{CP} allows for individual-alternative specific fixed effects. On the one hand, if the IIA property held, then both f 's would give rise to consistent estimators $\hat{\theta}_{FPH}$ and $\hat{\theta}_{CP}$, but result (2b) tells us that $\hat{\theta}_{FPH}$ would be more efficient than $\hat{\theta}_{CP}$. On the other hand, if the IIA property were violated in ways encompassed by individual-alternative specific fixed effects, then only $\hat{\theta}_{CP}$ would be consistent. It follows that, under the maintained assumption of choice set stability in $1, 2, \dots, T$, a test for H_0 : (IIA property) is $LR = 2 [l_{CP}(\hat{\theta}_{CP}) + l_{\Delta}(\hat{\theta}_{\Delta}) - l_{FPH}(\hat{\theta}_{FPH})]$.

F.1.2. Comparisons of same f with different choice sub-sequences

It is always possible to split choice sequences of length $1, 2, \dots, T^L$ into two (or more) *mutually exclusive* sub-sequences $1, 2, \dots, T^Z$ and $T^Z + 1, \dots, T^L$. Then $f_{CP}^Z(Y_i) \subset f_{CP}^L(Y_i)$ for any $Y_i \in \mathcal{CS}_i^* = c$. The same holds also for f_{FPH} and f_{PPH} . This method of making comparisons allows one to test for choice set stability in several alternative ways, but it does not enable one to test for departures from the IIA property (the two SSL models compared are always either both satisfying or both violating the IIA property).

Choice set stability: f_{CP} example. In what follows we will show with an example that $f_{CP}^Z(Y_i) \subset f_{CP}^L(Y_i)$ for any $Y_i \in \mathcal{CS}_i^* = c$ and afterward we will discuss how to use this fact to construct tests of choice set stability.

Suppose $J = 5$, $T^L = 4$, and that individual i is observed to make the choice sequence $Y_i = (j_1, j_2, j_3, j_4) = (3, 5, 5, 4)$.⁴³ By considering the observed choice sequence "at once", $Y_i = (3, 5, 5, 4)$ can be re-ordered in 12 different choice sequences.⁴⁴ Collect these sequences into the set $f_{CP}^L(Y_i) = l$. Assume that $V_i(X_{ijt}, \theta) = \delta_{ijt} + X_{ijt}\beta$. Then, i 's likelihood contribution given $f_{CP}^L(Y_i) = l$ is:

$$\begin{aligned} & \Pr[Y_i = (3, 5, 5, 4) | f_{CP}^L(Y_i) = l, \beta] \\ &= \frac{\exp((X_{i31} + X_{i52} + X_{i53} + X_{i44})\beta)}{\sum_{(j_1, j_2, j_3, j_4) \in f_{CP}^L(Y_i) = l} \exp((X_{ij_11} + X_{ij_22} + X_{ij_33} + X_{ij_44})\beta)}. \end{aligned} \quad (F.1)$$

Differently, by splitting i 's observed choice sequence into two mutually exclusive pairs of choices $Y_{i1} = (3, 5)$ and $Y_{i3} = (5, 4)$, we get $f_{CP}^Z(Y_i) = f_{CP}^Z(Y_{i1}) \times f_{CP}^Z(Y_{i3})$ where $f_{CP}^Z(Y_{i1}) = \{(3, 5), (5, 3)\}$ and $f_{CP}^Z(Y_{i3}) = \{(5, 4), (4, 5)\}$. Then, i 's likelihood contribution given $f_{CP}^Z(Y_{i1}) = z_1$ and $f_{CP}^Z(Y_{i3}) = z_3$ is:

$$\begin{aligned} & \Pr[Y_i = (3, 5, 5, 4) | f_{CP}^Z(Y_i) = z_1 \times z_3, \beta] \\ &= \frac{\exp((X_{i31} + X_{i52})\beta)}{\exp((X_{i31} + X_{i52})\beta) + \exp((X_{i51} + X_{i32})\beta)} \\ & \times \frac{\exp((X_{i53} + X_{i44})\beta)}{\exp((X_{i53} + X_{i44})\beta) + \exp((X_{i43} + X_{i54})\beta)}. \end{aligned} \quad (F.2)$$

⁴³ Alternative three in the first choice situation, alternative five in the second choice situation, etc.

⁴⁴ These sequences are: (3, 5, 5, 4), (5, 3, 5, 4), (5, 5, 3, 4), (5, 5, 4, 3), (4, 3, 5, 5), (3, 4, 5, 5), (3, 5, 4, 5), (5, 3, 4, 5), (5, 4, 3, 5), (5, 4, 5, 3), (4, 5, 3, 5), and (4, 5, 5, 3).

By multiplying the binomial logits in (F.2), we get:

$$\Pr_i [Y_i = (3, 5, 5, 4) | f_{CP}^Z(Y_i) = z, \beta] = \frac{\exp((X_{i31} + X_{i52} + X_{i53} + X_{i44})\beta)}{\sum_{(j_1, j_2, j_3, j_4) \in f_{CP}^Z(Y_i)=z} \exp((X_{ij_11} + X_{ij_22} + X_{ij_33} + X_{ij_44})\beta)}, \quad (F.3)$$

where $f_{CP}^Z(Y_i) = z$ collects sequences: (3, 5, 5, 4), (3, 5, 4, 5), (5, 3, 5, 4), and (5, 3, 4, 5). Consequently $f_{CP}^Z(Y_i) = z \subseteq f_{CP}^L(Y_i) = l$. In this example, f_{CP}^Z only uses information about 4 of the 12 possible choice sequences in f_{CP}^L . This implies that if unobserved choice sets were stable, then estimator $\hat{\beta}_{CP}^Z$ would be more efficient than $\hat{\beta}_{CP}^L$.

Moreover, the CP SSL estimated on choice sub-sequences may “discard” some choice situations: in the current example of sub-sequences of length two, whenever $j_t = j_{t+1}$ in $Y_{it} = (j_t, j_{t+1})$, then “fragment” Y_{it} of Y_i will not be used in estimation. For example, if i were observed to choose the sequence $Y_i = (3, 4, 5, 5)$, then only $Y_{i1} = (3, 4)$ would contribute to the likelihood function $l_{CP}^Z(\beta)$, while $l_{CP}^L(\beta)$ would still use the whole sequence $Y_i = (3, 4, 5, 5)$. More precisely, if $Y_i = (3, 4, 5, 5)$ were observed, then $f_{CP}^L(3, 4, 5, 5) = f_{CP}^L(3, 5, 5, 4) = l$ would still contain the same 12 choice sequences, while model (F.2) would collapse to:

$$\begin{aligned} \Pr [Y_i = (3, 4, 5, 5) | f_{CP}^Z(Y_i) = h_1 \times h_3, \beta] \\ &= \frac{\exp((X_{i31} + X_{i42})\beta)}{\exp((X_{i31} + X_{i42})\beta) + \exp((X_{i41} + X_{i32})\beta)} \\ &\times \frac{\exp((X_{i53} + X_{i54})\beta)}{\exp((X_{i53} + X_{i54})\beta)} \\ &= \frac{\exp((X_{i31} + X_{i42} + X_{i53} + X_{i54})\beta)}{\exp((X_{i31} + X_{i42} + X_{i53} + X_{i54})\beta) + \exp((X_{i41} + X_{i32} + X_{i53} + X_{i54})\beta)} \\ &= \Pr [Y_i = (3, 4, 5, 5) | f_{CP}^Z(Y_i) = h, \beta], \end{aligned} \quad (F.4)$$

which is also equivalent to $\Pr [Y_{i1} = (3, 4) | f_{CP}^Z(Y_{i1}) = h_1, \beta]$. In this case, then, $f_{CP}^Z(Y_{i1}) = h_1 \subset f_{CP}^Z(Y_i) = z \subset f_{CP}^L(Y_i) = l$. By result (2b), we can rank the corresponding estimators in terms of their relative efficiency. As a consequence, by splitting up choice sequences into mutually exclusive sub-sequences, one can face also this further loss of efficiency.

Model (F.1) requires stronger assumptions than model (F.3) for its consistent estimation. Consistent estimation of model (F.1) requires that alternatives $\{3, 4, 5\} \subseteq CS_{it}^* = c_t$, $t = 1, 2, 3, 4$. However, consistent estimation of model (F.3) only requires that $\{3, 5\} \subseteq CS_{it}^* = c_t$, $t = 1, 2$ and that $\{4, 5\} \subseteq CS_{it}^* = c_t$, $t = 3, 4$. In this example, if $4 \notin CS_{it}^* = c_t$, $t = 1$ or 2 , or $3 \notin CS_{it}^* = c_t$, $t = 3$ or 4 , then estimation of model (F.1) would not be consistent, while estimation of model (F.3) would.

These differences in consistency and relative efficiency suggest a Hausman test for unobserved choice set stability. If $\{3, 4, 5\} \subseteq CS_{it}^* = c_t$, $t = 1, 2, 3, 4$, then estimation of both model (F.1) and model (F.3) would be consistent. However, estimation of model (F.1) would be more efficient than estimation of model (F.3). If $4 \notin CS_{it}^* = c_t$, $t = 1$ or 2 or $3 \notin CS_{it}^* = c_t$, $t = 3$ or 4 , then only estimation of model (F.3) would be consistent. It follows that, under the maintained assumption of unobserved preference heterogeneity in a form encompassed by individual-alternative specific fixed effects, a test for H_0 : (choice set stability in 1, 2, 3, and 4) is $LR = 2 [l_{CP}^Z(\hat{\beta}_{PPH}^Z) + l_{\Delta}(\hat{\beta}_{\Delta}) - l_{CP}^L(\hat{\beta}_{CP}^L)]$.

Appendix G. Specification tests: Nested logit and IIA

In this Appendix, we illustrate that with similar assumptions to those required by the MNL, sufficient sets can also be used for the consistent estimation of the *within-nest* part of a nested logit model when choice sets are unobserved, and that this is enough to implement a test for departures of the IIA along the lines of Hausman and McFadden (1984).

Suppose that the full collection of J alternatives is partitioned into N mutually exclusive nests $nest_n$ and that any individual i 's choice set CS_{it}^* can be partitioned in N subsets of the N original nests, so that: $CS_{it}^* = nest_{i1} \cup \dots \cup nest_{in} \cup \dots \cup nest_{iN}$, where for any n , $nest_{in}$ is either $nest_{in} \subseteq nest_n$ or empty. The econometrician knows $nest_n$, $n = 1, \dots, N$, but does not know $nest_{in}$, $n = 1, \dots, N$, for any i . Note that, for simplicity, we are assuming that both the original nests and individual i 's nest subsets are constant over t . At the expense of some additional notation, this can be relaxed as in the case of the MNL. Denote by $Y_i = (Y_{i1}, \dots, Y_{iT})$ individual i 's sequence of chosen alternatives and by $Q_i^* = (Q_{i1}^*, \dots, Q_{iT}^*)$ i 's sequence of chosen nests, with $Y_{it} \in Q_{it}^*$ and $Q_{it}^* \in \{nest_{i1}, \dots, nest_{iN}\}$. Define $Q_i^* = \times_{t=1}^T Q_{it}^*$ and note that $Q_i^* \subseteq CS_{it}^* = \times_{t=1}^T CS_{it}^*$, so that for each t one has $Q_{it}^* \subseteq CS_{it}^*$. The econometrician observes the realization of Y_i and knows to which of the original nests each Y_{it} belongs, for example $Y_{it} \in nest_n$, but does not know much about Q_{it}^* beyond the facts that $Y_{it} \in Q_{it}^*$ and that $Q_{it}^* = nest_{in} \subseteq nest_n$.

Assumption 4. Conditional on all $V(X_{ijt}, \theta) = X_{ijt}\theta$'s and on $CS_i^* = c$, $\Pr[Y_i = j, Q_i^* = q | CS_i^* = c, \theta, \lambda] = \prod_{t=1}^T \Pr[Y_{it} = j_t, Q_{it}^* = q_t | CS_{it}^* = c_t, \theta, \lambda]$ is a product of T per-period nested logits as in Eq. (G.1) with $\lambda = (\lambda_1, \dots, \lambda_N)$ being the nesting parameters associated to each nest.

Suppose that $Y_i = j$, $Q_i^* = q$, and $CS_i^* = c$. The nested logit model can be expressed as:

$$\begin{aligned} \Pr[Y_i = j, Q_i^* = q | CS_i^* = c, \theta, \lambda] &= \prod_{t=1}^T \overbrace{\Pr[Y_{it} = j_t, Q_{it}^* = q_t | CS_{it}^* = c_t, \theta, \lambda]}^{\text{per-period nested logit}} \\ &= \prod_{t=1}^T \Pr[Y_{it} = j_t | CS_{it}^* = c_t, Q_{it}^* = q_t, \theta, \lambda] \times \Pr[Q_{it}^* = q_t | CS_{it}^* = c_t, \theta, \lambda] \\ &= \prod_{t=1}^T \overbrace{\Pr[Y_{it} = j_t | Q_{it}^* = q_t, \theta, \lambda]}^{\text{within-nest MNL}} \times \underbrace{\prod_{t=1}^T \Pr[Q_{it}^* = q_t | CS_{it}^* = c_t, \theta, \lambda]}_{\text{between-nest MNL}} \\ &= \prod_{t=1}^T \frac{\exp(X_{ijt}\theta/\lambda_{q_t})}{\sum_{v_t \in q_t} \exp(X_{iv_t}\theta/\lambda_{q_t})} \times \prod_{t=1}^T \Pr[Q_{it}^* = q_t | CS_{it}^* = c_t, \theta, \lambda], \end{aligned} \quad (\text{G.1})$$

which is a function of the unobserved realizations q and c . When all the nesting parameters equal one, $(\lambda_1, \dots, \lambda_N) = 1$, then the nested logit in (G.1) simplifies to a standard MNL. In order to test for this hypothesis, it is enough to obtain a consistent estimator of the $(\theta/\lambda_n)_{n=1}^N$ parameters of the *within-nest* MNL model:

$$\begin{aligned} \Pr[Y_i = j | Q_i^* = \times_{t=1}^T q_t, \theta, \lambda] &= \prod_{t=1}^T \Pr[Y_{it} = j_t | Q_{it}^* = q_t, \theta, \lambda] \\ &= \prod_{t=1}^T \frac{\exp(X_{ijt}\theta/\lambda_{q_t})}{\sum_{v_t \in q_t} \exp(X_{iv_t}\theta/\lambda_{q_t})} \end{aligned} \quad (\text{G.2})$$

and check whether $\theta/\lambda_m = \theta/\lambda_n$ for all $m \neq n$.⁴⁵ We define as a sufficient set for the within-nest MNL any correspondence that satisfies the following condition.

Condition 3. Given any choice sequence $Y_i \in Q_i^* \subseteq CS_i^*$, the correspondence f is such that $Y_i \in f(Y_i)$ and $f(Y_i) \subseteq Q_i^*$, with $f(Y_i) = \times_{t=1}^T f_t(Y_i)$ and each $f_t(Y_i)$ so that $Y_{it} \in f_t(Y_i) \subseteq Q_{it}^*$.

In words, given any sequence of choices Y_i , a sufficient set f enables the econometrician to define a corresponding sequence of nest subsets $f(Y_i) = \times_{t=1}^T f_t(Y_i)$, where each $f_t(Y_i)$ is a subset of the specific *nest_{in}* chosen by i in t . Given Assumption 4 and Condition 3, the within-nest MNL from (G.2) conditional on $f(Y_i) = r$ simplifies to:

$$\begin{aligned} \Pr[Y_i = j | Q_i^* = \times_{t=1}^T q_t, f(Y_i) = r, \theta, \lambda] &= \Pr[Y_i = j | f(Y_i) = r, \theta, \lambda], \text{ because } r \subseteq \times_{t=1}^T q_t \\ &= \prod_{t=1}^T \Pr[Y_{it} = j_t | f_t(Y_i) = r_t, \theta, \lambda], \text{ because of Eq. (3.6)} \\ &= \prod_{t=1}^T \frac{\exp(X_{ijt}\theta/\lambda_{r_t})}{\sum_{v_t \in r_t} \exp(X_{iv_t}\theta/\lambda_{r_t})}, \end{aligned} \quad (\text{G.3})$$

which, a part from the parameters of interest, only depends on observed quantities. Given the assumption of choice set stability, a sufficient set compatible with Condition 3 is the Within(-nest) Full Purchase History, or WFPH. This is similar to the FPH sufficient set, but now one should separately keep track of the alternatives purchased by i over choice situations within each of the N nests. Define the set of alternatives ever purchased by i in *nest_n* by $H_i^n = \cup_{t=1}^T \{Y_{it} | Y_{it} \in \text{nest}_n\}$ for

⁴⁵ The first equality in (G.2) follows from Eq. (3.6), because $Q_i^* = \times_{t=1}^T q_t$.

$n = 1, \dots, N$. Note that, for any $Y_{it} \in \text{nest}_{in}$, $H_i^n \subseteq Q_{it}^* = \text{nest}_{in}$. Then any observed sequence of chosen alternatives Y_i will enable one to construct the sufficient set $f_{WFFPH}(Y_i) = \times_{t=1}^T \{H_i^n | Y_{it} \in \text{nest}_n\}$.

The idea of the IIA test is then simple. For a given sufficient set f from Condition 3, one can estimate a variant of the SSL with a different $\theta_n = \theta/\lambda_n$ for each nest $n = 1, \dots, N$. Similarly, one can estimate N separate SSL models, each from the observed choices within each of the N nests. Then, if the estimated $\theta_m \neq \theta_n$ for at least two nests $n \neq m$, the econometrician will have evidence of violations of the IIA property. The validity of this testing procedure, similar to the original one proposed by Hausman and McFadden (1984), rests on the maintained assumption that f is a valid sufficient set.

Appendix H. Data appendix

In this Appendix, we describe in greater detail the data used in the empirical illustration in Section 5.

H.1. Purchase data

We use data from the Kantar Worldpanel (see Leicester and Oldfield, 2009, Dubois et al., 2018, and Dubois et al., 2020). Kantar collects data on purchases made on-the-go from a random selection of individuals in the households that participate in the Worldpanel. The Kantar Worldpanel on-the-go survey is collected from individuals who record purchases that they make on-the-go for immediate consumption using their mobile phone.

H.2. Advertising data

To measure advertising exposure, we convert weekly advertising (“flows”) into an advertising “stock;” advertising stocks are the depreciated accumulation of the flows. We use advertising data collected by AC Nielsen on TV advertising. TV advertising accounts for 61.8% of total expenditure on chocolate bar advertising over this period.

For each TV ad, we have information on the time the ad was aired, the brand that was advertised, the TV station, the duration of the ad, the cost of the ad, and the TV shows that immediately preceded and followed the ad. The time path of advertising varies across brands, and all brands have some periods of zero advertising expenditure. These non-smooth strategies are rationalized in the model of Dubé et al. (2005) when the effectiveness of advertising can vary over time. This variation in the timing of adverts, coupled with variation in TV viewing behavior, generates household level variation in exposure to brand level advertising.

Our advertising measure follows Goeree (2008) and Dubois et al. (2018) and measures advertising exposure at the individual level. We use detailed information about when individual adverts were aired on television matched with self-reported viewing information to construct individual level measures of exposure to brand advertising. We use data from the Kantar media survey, an annual survey asking the main shopper in the household about their TV subscriptions and TV viewing behavior. Households are asked “How often do you watch ...?” for 206 different TV shows, and can choose to answer Never, Hardly Ever, Sometimes or Regularly. At least one ad for chocolate is shown before, during, or after 112 of these shows (many of the shows with no chocolate advertising are on BBC channels, which are prohibited from showing ads). From this information we define the variable:

$$w_{is} = \begin{cases} 1 & i \text{ reports they “regularly” or “sometimes” watch show } s \\ 0 & \text{otherwise} \end{cases} \quad (\text{H.1})$$

Households are also asked “How often do you watch ...?” 65 different TV channels and when they usually watch TV. In particular, for weekdays, Saturday, and Sunday and for 9 different time periods,⁴⁶ households are asked questions like “Do you watch live TV on Saturdays at breakfast time (6.00–9.30 am)?” In each case, the household can answer Never, Hardly Ever, Sometimes or Regularly. We use this information, along with information on where the household lives (some TV channels are regional), to construct the variable:

$$w_{ikc} = \begin{cases} 1 & i \text{ says they “regularly” or “sometimes” watch on the day and time slot } k \\ & \text{and “regularly” or “sometimes” watch channel } c \\ & \text{and they live in the region in which } c \text{ is aired (or the channel is national)} \\ 0 & \text{otherwise} \end{cases} \quad (\text{H.2})$$

We combine the data on household viewing behavior with the detailed data on individual ads to create a household-specific measure of exposure to advertising. Variation in TV viewing behavior creates considerable variation in the timing and extent of exposure an individual household has to ads of a specific brand. This leads to cross-household variation in advertising exposure that is plausibly unrelated to idiosyncratic shocks to demand for chocolate products.

Denote by T_{bskt} the duration of time that an ad for brand b is shown during show s on day and time slot k on channel c during week t . From the viewing data, we construct an indicator variable of whether household i was likely to be watching

⁴⁶ Breakfast time 6.00 am–9.30 am, Morning 9.30 am–12.00 noon, Lunchtime 12.00 noon–2.00 pm, Early afternoon 2.00 pm–4.00 pm, Late afternoon 4.00 pm–6.00 pm, Early evening 6.00pm–8.00pm, Mid evening 8.00 pm–10.30 pm, Late evening 10.30–1.00 am and Night time 1.00 am–6.00 am.

channel c on day and time slot k during show s , w_{iskc} . If show s is among the 206 specific shows households were asked for viewing information we set $w_{iskc} = w_{is}$, otherwise we set $w_{iskc} = w_{ikc}$. From this we define the household's total exposure to advertising of brand b during week t (their weekly advertising "flow") as:

$$s_{ibt} = \sum_{s,c,k} w_{iskc} T_{bskct}. \quad (\text{H.3})$$

We define a household's accumulated advertising *stock* to brand b in week t as the depreciated accumulation of these advertising flows:

$$stock_{ibt} = \sum_{k=0}^t \eta^k s_{ibt-k} \quad (\text{H.4})$$

where $\eta = 0.75$. This stock is measured in seconds (and is divided by 1000 when included in the regression). It is 0 for individuals that do not watch TV, or only watch public TV (the BBC), and has a mean of 10 min of cumulated exposure to adverts for a particular brand.

Finally, we follow Dubé et al. (2005) and allow for diminishing returns to advertising by transforming the stock of advertising, $stock_{ibt}$, using the log inverse hyperbolic sine function, $\ln(a_{ibt}) = \ln\left(stock_{ibt} + \sqrt{1 + stock_{ibt}^2}\right)$.

References

- Abaluck, Jason, Adams, Abi, 2017. What do consumers consider before they choose: Identification from asymmetric demand responses. NBER Working Paper 23566.
- Arellano, Manuel, 2003. Panel Data Econometrics. Oxford University Press.
- Bajari, Patrick, Fox, Jeremy T., Ryan, Stephen P., 2007. Linear regression estimation of discrete choice models with nonparametric distributions of random coefficients. *Amer. Econ. Rev.* 97 (2), 459–463.
- Bajari, Patrick, Fox, Jeremy T., Ryan, Stephen P., 2008. Evaluating wireless carrier consolidation using semiparametric demand estimation. *Quant. Mark. Econ.* 6 (4), 299.
- Barseghyan, Levon, Coughlin, Maura, Molinari, Francesca, Teitelbaum, Joshua C., 2019. Heterogeneous choice sets and preferences. arXiv preprint arXiv:1907.02337.
- Başar, Gözen, Bhat, Chandra, 2004. A parameterized consideration set model for airport choice: an application to the San Francisco Bay area. *Transp. Res. B* 38 (10), 889–904.
- Becker, G.S., Murphy, K.M., 1993. A simple theory of advertising as a good or ban. *Q. J. Econ.* 108 (4), 941–964.
- Ben-Akiva, M., Boccara, Brian, 1995. Discrete choice models with latent choice sets. *Int. J. Res. Mark.* 9–24.
- Berry, S., 1994. Estimating discrete choice models of product differentiation. *Rand. J. Econ.* 25 (2), 242–262.
- Berry, Steven, Levinsohn, James, Pakes, Ariel, 2004. Differentiated products demand systems from a combination of micro and macro data: The new car market. *J. Political Econ.* 68–105.
- Bierlaire, Michel, Bolduc, Denis, McFadden, Daniel, 2008. The estimation of generalized extreme value models from choice-based samples. *Transp. Res. B* 42 (4), 381–394.
- Bordalo, P., Gennaioli, N., Shleifer, A., 2014. Salience and consumer choice. *J. Political Econ.* 121 (5), 803–843.
- Bronnenberg, B., Vanhonacker, V., 1996. Limited choice sets, local price response, and implied measures of price competition. *J. Market. Res.* 163–173.
- Bruno, Hernán A., Vilcassim, Naufel J., 2008. Research note—Structural demand estimation with varying product availability. *Mark. Sci.* 27 (6), 1126–1131.
- Caplin, Andrew, Dean, Mark, 2011. Search, choice, and revealed preference. *Theor. Econ.* 6 (1), 19–48.
- Caplin, Andrew, Dean, Mark, Martin, Daniel, 2011. Search and satisficing. *Am. Econ. Rev.* 101 (7), 2899–2922.
- Chamberlain, Gary, 1980. Analysis of covariance with qualitative data. *Rev. Econom. Stud.* 47, 225–238.
- Chiang, Jeongwen, Chib, Siddhartha, Narasimhan, Chakravarthi, 1998. Markov chain Monte Carlo and models of consideration set and parameter heterogeneity. *J. Econometrics* 89 (1), 223–248.
- Ching, Andrew T., Erdem, Tülin, Keane, Michael P., 2014. A simple method to estimate the roles of learning, inventories and category consideration in consumer choice. *J. Choice Model.* 13, 60–72.
- Ciliberto, Federico, Murry, Charles, Tamer, Elie T., 2016. Market structure and competition in airline markets. Working paper.
- Conlon, Christopher T., Mortimer, Julie Holland, 2013. Demand estimation under incomplete product availability. *Am. Econ. J.-Microecon.* 5 (4), 1–30.
- Currie, Janet, DellaVigna, Stefano, Moretti, Enrico, Pathania, Vikram, 2010. The effect of fast food restaurants on obesity and weight gain. *Am. Econ. J.: Econ. Policy* (ISSN: 1945-7731, 1945-7731) 2 (3), 32–63. <http://dx.doi.org/10.1257/pol.2.3.32>.
- De los Santos, Babur, Hortaçsu, Ali, Wildenbeest, Matthijs R., 2012. Testing models of consumer search using data on web browsing and purchasing behavior. *Am. Econ. Rev.* 102 (6), 2955–2980.
- de Palma, André, Picard, Nathalie, Waddell, Paul, 2007. Discrete choice models with capacity constraints: An empirical analysis of the housing market of the greater Paris region. *J. Urban Econ.* 204–230.
- D'Haultfoeulle, Xavier, Iaria, Alessandro, 2016. A convenient method for the estimation of the multinomial logit model with fixed effects. *Econom. Lett.* 141, 77–79.
- Draganska, Michaela, Klapper, Daniel, 2011. Choice set heterogeneity and the role of advertising: An analysis with micro and macro data. *J. Market. Res.* 48 (4), 653–669.
- Draganska, M., Mazzeo, M., Seim, K., 2009. Beyond plain vanilla: Modeling joint product assortment and pricing decisions. *Quant. Mark. Econ.* 7, 105–146.
- Dubé, J.P., Hitsch, G., Manchanda, P., 2005. An empirical model of advertising dynamics. *Quant. Mark. Econ.* 3, 107–144.
- Dubois, P., Griffith, R., O'Connell, M., 2018. The effects of banning advertising in junk food markets. *Rev. Econ. Stud.* 85 (1), 396–436.
- Dubois, P., Griffith, R., O'Connell, M., 2020. How well targeted are soda taxes?. *Am. Econ. Rev.* forthcoming.
- Eizenberg, A., 2014. Upstream innovation and product variety in the U.S. home PC market. *Rev. Econom. Stud.* 81, 1003–1045.
- Eliaz, Kfir, Spiegel, Ran, 2011. Consideration sets and competitive marketing. *Rev. Econom. Stud.* 78 (1), 235–262.
- Erdem, Tülin, Swait, Joffre, 2004. Brand credibility, brand consideration, and choice. *J. Consum. Res.* 31 (1), 191–198.

- Fack, G., Grenet, J., He, Y., 2019. Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *Am. Econ. Rev.* 109 (4), 1486–1529.
- Fox, Jeremy T., 2007. Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *Rand J. Econ.* 38 (4), 1002–1019.
- Fox, Jeremy, il Kim, Kyoo, Ryan, Stephen, Bajari, Patrick, 2011. A simple estimator for the distribution of random coefficients. *Quant. Econ.* 2 (3), 381–418.
- Fox, Jeremy T., il Kim, Kyoo, Yang, Chenyu, 2016. A simple nonparametric approach to estimating the distribution of random coefficients in structural models. *J. Econometrics* 195 (2), 236–254.
- Frejinger, Emma, Bierlaire, Michel, Ben-Akiva, Moshe, 2009. Sampling of alternatives for route choice modeling. *Transp. Res. B* 43 (10), 984–994.
- Freyberger, Joachim, 2015. Asymptotic theory for differentiated products demand models with many markets. *J. Econometrics* 185 (1), 162–181.
- Gabaix, Xavier, 2014. A sparsity-based model of bounded rationality. *Q. J. Econ.* 129 (4), 1661–1710.
- Gaynor, Martin, Propper, Carol, Seiler, Stephan, 2016. Free to choose? Reform, choice and consideration sets in the English national health service. *Amer. Econ. Rev.* forthcoming.
- Gilbride, Timothy J., Allenby, Greg M., 2004. A choice model with conjunctive, disjunctive, and compensatory screening rules. *Mark. Sci.* 23 (3), 391–406.
- Goeree, Michelle Sovinsky, 2008. Limited information and advertising in the U.S. personal computer industry. *Econometrica* 76 (5), 1017–1074.
- Goeree, J.K., Holt, C.A., Palfrey, T.R., 2005. Regular quantal response equilibrium. *Exp. Econ.* 8 (4), 347–367.
- Guevara, C. Angelo, Ben-Akiva, Moshe E., 2013a. Sampling of alternatives in logit mixture models. *Transp. Res. B* 58, 185–198.
- Guevara, C. Angelo, Ben-Akiva, Moshe E., 2013b. Sampling of alternatives in multivariate extreme value (MEV) models. *Transp. Res. B* 48, 31–52.
- Hausman, Jerry, McFadden, Daniel, 1984. Specification tests for the multinomial logit model. *Econometrica* 1219–1240.
- Hausman, Jerry A., Ruud, Paul A., 1987. Specifying and testing econometric models for rank-ordered data. *J. Econometrics* 34 (1), 83–104.
- Heiss, Florian, McFadden, Daniel, Winter, Joachim, Wuppermann, Amelie, Zhou, Bo, 2016. Inattention and switching costs as sources of inertia in medicare part d. National Bureau of Economic Research. Working paper.
- Hey, John D., Orme, Chris, 1994. Investigating generalizations of expected utility theory using experimental data. *Econometrica* 1291–1326.
- Hickman, William, Mortimer, Julie Holland, 2016. Demand estimation with availability variation. In: *Handbook on the Economics of Retailing and Distribution*. pp. 306–342.
- Ho, K., Hogan, J., Scott Morton, F., 2017. The impact of consumer inattention on insurer pricing in the Medicare Part D program. *RAND J. Econ.* 48 (4), 877–905.
- Honka, Elisabeth, 2014. Quantifying search and switching costs in the US auto insurance industry. *Rand J. Econ.* 45 (4), 847–884.
- Honka, E., Hortaçsu, A., Wildenbeest, M., 2019. Empirical search and consideration sets. In: Dube, J-P., Rossi, P. (Eds.), Chapter 4 in *Handbook for the Economics of Marketing*, Volume 1. Elsevier, pp. 193–257.
- Hortaçsu, A., Madanizadeh, S., Puller, S., 2017. Power to choose? an analysis of consumer inertia in the residential electricity market. *Am. Econ. J.: Econ. Policy* 9 (4), 192–226.
- Iaria, Alessandro, 2014. Consumer benefit of big-box supermarkets: The importance of controlling for endogenous entry. Working paper.
- Imbens, Guido W., Manski, Charles F., 2004. Confidence intervals for partially identified parameters. *Econometrica* 1845–1857.
- Janssen, Maarten C.W., Moraga-González, José Luis, 2004. Strategic pricing, consumer search and the number of firms. *Rev. Econom. Stud.* 71 (4), 1089–1118.
- Juang, Y., Bronnenberg, B., 2018. Pennies for your thoughts: Costly product consideration and purchase quantity thresholds. *Mark. Sci.* 37 (6), 855–1052.
- Keane, Michael P., Wasi, Nada, 2012. Estimation of discrete choice models with many alternatives using random subsets of the full choice set: With an application to demand for frozen pizza. Working paper.
- Köszegi, Botond, Szeidl, Adam, 2013. A model of focusing in economic choice. *Q. J. Econ.* 128 (1), 53–104.
- Leicester, Andrew, Oldfield, Zoe, 2009. An analysis of consumer panel data, IFS Working Papers W09/09, 2009, Institute for Fiscal Studies.
- Li, Sophia Ying, Mazur, Joe, Park, Yongjoon, Roberts, James W., Sweeting, Andrew, Zhang, Jun, 2018. Endogenous and selective service choices after airline mergers, National Bureau of Economic Research.
- Lu, Zhentong, 2018. Estimating Multinomial Choice Models with Unobserved Choice Sets. Working paper.
- Manski, Charles F., 1975. Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* 3 (3), 205–228.
- Manski, Charles F., 1977. The structure of random utility models. *Theory and Decision* 8 (3), 229–254.
- Manzini, Paola, Mariotti, Marco, 2014. Stochastic choice and consideration sets. *Econometrica* 82 (3), 1153–1176.
- Masatlioglu, Yusufcan, Nakajima, Daisuke, 2013. Choice by iterative search. *Theor. Econ.* 8 (3), 701–728.
- Masatlioglu, Yusufcan, Nakajima, Daisuke, Ozbay, Erkut Y., 2012. Revealed attention. *Amer. Econ. Rev.* 2183–2205.
- Matejka, Filip, McKay, Alisdair, 2015. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *Amer. Econ. Rev.* 105 (1), 272–298.
- McFadden, Daniel, 1978. Modeling the choice of residential location. In: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. (Eds.), *Spatial Interaction Theory and Planning Models*, Vol. 1. North-Holland, pp. 75–96.
- McFadden, Daniel, 1987. Regression-based specification tests for the multinomial logit model. *J. Econometrics* 34 (1–2), 63–82.
- McFadden, Daniel, Tye, William B., Train, Kenneth, 1977. An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model. Institute of Transportation Studies, University of California.
- Morgan, Peter, Manning, Richard, 1985. Optimal search. *Econometrica* 923–944.
- Musalem, Andrés, 2015. When demand projections are too optimistic: A structural model of product line and pricing decisions. Working paper.
- Rhodes, Andrew, 2014. Multiproduct retailing. *Rev. Econom. Stud.* 82 (1), 360–390.
- Roberts, John H., Lattin, James M., 1991. Development and testing of a model of consideration set composition. *J. Market. Res.* 429–440.
- Ruud, Paul A., 1984. Tests of specification in econometrics. *Econometric Rev.* 3 (2), 211–242.
- Simon, Herbert A., 1955. A behavioral model of rational choice. *Q. J. Econ.* 69 (1), 99–118.
- Train, Kenneth E., McFadden, Daniel L., Ben-Akiva, Moshe, 1987. The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *Rand J. Econ.* 109–123.
- Train, K., Winston, C., 2007. Vehicle choice behavior and the declining market share of US automakers. *Int. Econ. Rev.*
- Van Nierop, Erjen, Bronnenberg, Bart, Paap, Richard, Wedel, Michel, Franses, Philip Hans, 2010. Retrieving unobserved consideration sets from household panel data. *J. Market. Res.* 47 (1), 63–74.
- Walters, C., 2018. The demand for effective charter schools. *Q. J. Econ.* 126 (6), 2179–2223.